Some new methods for three-mode factor analysis and multi-set factor analysis

Lam, Tam

SOME NEW METHODS FOR THREE-MODE FACTOR ANALYSIS AND MULTI-SET FACTOR ANALYSIS

LAM THI THANH TAM

university of groningen

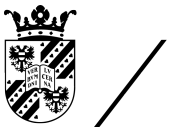
This PhD project was carried out at the Psychometrics and Statistics Group at the Heymans Institute for Psychological Research at the University of Groningen, according the requirements of the Graduate School of Behavioural and Social Sciences.

The funding of this research was provided by The Netherlands Organisation for Scientific Research (NWO), VIDI Research Grant 452-08-001.

Printed by Ipskamp Drukkers, Enschede, The Netherlands.

ISBN: 978-90-367-7544-1 (print version)

ISBN: 978-90-367-7543-4 (electronic version)



university of groningen

Some new methods for three-mode factor analysis and multi-set factor analysis

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. E. Sterken
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Thursday 19 February 2015 at 16.15 hours

by

Lam Thi Thanh Tam

born on 14 July 1980

in Quang Ngai, Vietnam

Supervisor

Prof. R.R. Meijer

Co-supervisor

Dr. A.W. Stegeman

Assessment committee

Prof. F.J. Oort

Prof. H.A.L. Kiers

Prof. J.M.F. ten Berge

to my dear family.

Contents

Chapter 1

Introduction

For summarizing the information contained in a dataset of observations on a number of correlated variables, traditionally the methods of Principal Component Analysis and Common Factor Analysis are used.

Principal Component Analysis (PCA) was introduced by Pearson (1901). This method is now used as a tool in exploratory data analysis and for making predictive models. It can convert a set of observations of correlated variables into a reduced set of values of uncorrelated variables called principal components. The first principal component captures the largest possible amount of the information contained in the original data, and each succeeding component in turn captures the highest possible amount of the remaining information contained in the original data under the constraint that it is uncorrelated with the preceding components. This allows to reduce the dimension of the original data set while losing the smallest possible amount of information.

Common Factor Analysis arose around the same period (with Spearman, 1904). Here, each variable is assumed to have a unique part which does not correlate with other variables. Common Factor Analysis is a correlation-focused

approach seeking to reproduce the inter-correlations among variables, in which the factors represent the common variance of variables, excluding unique variance.

In general, both PCA and Common Factor Analysis are techniques which seek to summarize observed variables into a smaller set of components/factors. However, they are typically used for different reasons, are not computationally the same, nor do they have the same underlying model. Common Factor Analysis takes into account the unique part of each variable while this is neglected in PCA. The unique part may be seen as measurement error or as a part of the variable that does not correlate with other variables or as a combination of both. The lively discussion of Common Factor Analysis versus PCA has been documented extensively in the literature; see e.g. Velicer and Jackson (1990) for an overview. PCA can be favored because of its computational simplicity and manifest component approach, whereas Common Factor Analysis is computationally more difficult and features latent factors.

Multi-way data refer to data that can be arranged in a multi-dimensional array or multi-way array. In this thesis, we consider three-way data which is found in many different contexts. For example: scores on various anxiety scales of a number of individuals in various situations; scores on various competences of a number of workers assessed by several different assessors (i.e. Multitrait-Multimethod data); scores on food quality indicators of a number of food products by several different judges; and fMRI brain scan measurements for different areas of the brain over a period of time for different individuals. The three sets of entities associated with such three-way data sets are called the three *modes* of the array.

Multi-set data refer to data when it consists of a set of data matrices and each data matrix has the same column units (i.e., variables) but different row

units (i.e., observational units). For example, we have multi-set data when the same variables are observed for several different populations or subpopulations.

In this thesis, we focus on Exploratory Component Analysis and Common Factor Analysis techniques for multi-way and multi-set data. The thesis is organized as follows.

In Chapter 2, we introduce the concepts and properties of multi-way and multi-set component and factor analysis whose comprehension is crucial to fully understand all the subsequent chapters of this thesis. These concepts and properties concern two-way and three-way decompositions (e.g. Singular Value Decomposition, Principal Component Analysis for two-way decomposition; Candecomp/Parafac, Tucker3 for three-way decomposition), multi-set component and factor analysis (e.g. Parafac2, Simultaneous Component Analysis models, Multi-set Parafac2 common factor model). Also, the uniqueness properties of the Candecomp/Parafac and Tucker3 models as well as algorithms to fit these models are discussed.

Chapter 3 discusses Exploratory Component Analysis of a 4-way dataset of Belief in a Just World (BJW) data. The subjects are asked how strongly they believe that a number of 6 actors (Natural, God, Human Institutions, Other People, Yourself, and Chance) bring about justice in the world for themselves or other people. The 4 modes of the dataset are: 345 subjects, 7 items, 6 actors, and 2 perspectives (For Yourself and For Others). Our analysis includes exploring the correlation structure and conducting a PCA of two-way unfoldings of (part of) the dataset. Also, we fit Tucker3 to three-way parts of the dataset, and Tucker4 to the complete dataset. We also discuss how to rotate a Tucker4 rotation to simple structure.

In Chapters 4, 5, 6, we propose and demonstrate new methods of three-mode and multi-set factor analysis that make use of Minimum Rank Factor Analysis

(MRFA), and Candecomp/Parafac or Tucker3 (for three-mode factor analysis; Chapters 4 and 5), and Parafac2 (for multi-set factor analysis; Chapter 6).

By using our methods in Chapters 4, 5, and 6, one can compute the overall percentage of explained common variance, and also for each variable separately. This is usually not possible for existing methods of three-mode or multi-set factor analysis. Moreover, the algorithms that we propose are simple and easy to run. Our solutions are easy to interpret and our models are parsimonious.

Solutions of the methods in Chapters 4 and 6 are rotationally unique due to uniqueness properties of Candecomp/Parafac and Parafac2. Solutions of the method in Chapter 5 are not unique due to non-uniqueness of Tucker3. Therefore, we use the Joint Orthomax rotation of (Kiers, 1998a) to obtain simple solutions in Chapter 5.

In Chapter 7 we classify the three-mode and multi-set component and factor models featuring in this thesis in a $2 \times 2 \times 3$ array of models. We briefly discuss how a suitable model can be chosen to analyze a three-mode or multi-set dataset. Also, we indicate topics for future research.

Notations

The following notation is used throughout this thesis.

- In general, scalar values are denoted with lower case italic font: a, x, y, ...; Column vectors are denoted with lower case bold-face font: $\mathbf{a}$, $\mathbf{x}$, $\mathbf{y}$, ...; Matrices are denoted with upper case bold-face font: $\mathbf{A}$, $\mathbf{X}$, $\mathbf{Y}$,...; Three-way arrays are denoted with underlined upper case bold-face font: $\underline{\mathbf{A}}$, $\underline{\mathbf{X}}$, $\underline{\mathbf{Y}}$, ...
- $\mathbf{A}^T$: The transpose of matrix $\mathbf{A}$.

- $\mathrm{tr}(\mathbf{A})$: The trace of matrix $\mathbf{A}$.
- $\mathrm{Vec}(\mathbf{A})$: The column vector that is formed by stacking the columns of matrix $\mathbf{A}$ on top of each other, with the first column on top.
- $\mathrm{Vec}(\underline{\mathbf{X}})$: The column vector is formed by stacking column vectors $\mathrm{Vec}\left(\mathbf{X}_1^{(12)}\right)$, $\mathrm{Vec}\left(\mathbf{X}_2^{(12)}\right)$, $\ldots$, $\mathrm{Vec}\left(\mathbf{X}_K^{(12)}\right)$.
- $\circ$: The outer product of vectors.
- $\odot$: The Khatri-Rao product of matrices.
- $\otimes$: The Kronecker product of matrices.
- $\mathrm{diag}(\mathbf{a})$: The diagonal matrix with the entries of vector $\mathbf{a}$ on its diagonal.

Chapter 2

Multi-way and multi-set component and factor models

During the last decades, the interest in multi-way data representations has increased exponentially. When data is described as N-way ($N > 2$), a parsimonious representation is often desirable. Such representations are given by multi-way generalizations of the Principal Component Analysis (PCA) or of the Singular Value Decomposition (SVD). They are called multi-way decompositions or tensor decompositions.

A multi-way decomposition is a decomposition of a multi-way array into a sum of rank-1 arrays. In a multi-way model, a multi-way data array is the sum of a multi-way decomposition and a residual array. The optimal solution of a multi-way model is usually the multi-way decomposition that is found by minimizing the sum-of-squares of the residual array.

Different from PCA, Factor Analysis is a correlation-focused approach seeking to reproduce the inter-correlations among variables, in which the factors represent the common variance of variables, excluding unique variance. Here,

each variable is assumed to have a unique part which does not correlate with other variables.

In section 2.1, we introduce the general framework of two-way and three-way decomposition. In section 2.2, three-way decomposition is introduced with existing models and methods (e.g. Candecomp/Parafac and Tucker3). In section 2.3, we discuss uniqueness conditions for the Candecomp/Parafac model. In section 2.4, we discuss non-existence of a best-fitting Candecomp/Parafac model. In section 2.5, we discuss two-way and three-way Common Factor Analysis. Finally, in section 2.6, we discuss multi-set component and factor analysis.

2.1 Two-way decomposition

The two-way decomposition is a decomposition of a matrix (two-way array) into a sum of rank-1 matrices. It comprises various bilinear methods (SVD, Factor Analysis, PCA) and has a strong relation to the rank of a matrix, which is defined as the smallest number of rank-1 matrices whose sum equals the matrix.

The model containing the two-way decomposition of a $I \times J$ matrix $\mathbf{X}$ is described by the following equation

$$\mathbf{X} = \mathbf{A}\mathbf{B}^T + \mathbf{E} = \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r + \mathbf{E}, \tag{2.1}$$

where $\mathbf{E}$ $(I \times J)$ is the residual matrix, and $\mathbf{A}$ $(I \times R)$ and $\mathbf{B}$ $(J \times R)$ are matrices whose columns are $\mathbf{a}_1, ..., \mathbf{a}_R$ and $\mathbf{b}_1, ..., \mathbf{b}_R$, respectively.

The notation "$\circ$" stands for the outer product of two vectors, defined as $\mathbf{x} \circ \mathbf{y} = \mathbf{x}\mathbf{y}^T$. And note that, since all columns of the matrix $\mathbf{x}\mathbf{y}^T$ are proportional, its rank is equal to 1 if $\mathbf{x}$ and $\mathbf{y}$ are non-zero. Hence, in (2.1), the matrix $\mathbf{X}$ is decomposed into the sum of R rank-1 matrices and a residual matrix $\mathbf{E}$. If $R = \text{rank}(\mathbf{X})$, that is $\mathbf{E} = \mathbf{O}$, then the pair $(\mathbf{A}, \mathbf{B})$ is called a perfect fitting

solution of (2.1) and $\mathbf{X} = \mathbf{A}\mathbf{B}^T$. If $R < \text{rank}(\mathbf{X})$, then $\mathbf{A}\mathbf{B}^T$ is called a rank-R approximation of $\mathbf{X}$. And, if $(\mathbf{A}, \mathbf{B})$ is a solution which minimizes $||\mathbf{E}||^2 = \sum_{i,j} e_{ij}^2$, then it is called an optimal solution of (2.1), and the corresponding $\mathbf{A}\mathbf{B}^T$ is called a best rank-R approximation of $\mathbf{X}$.

The equation (2.1) can be graphically depicted as follows

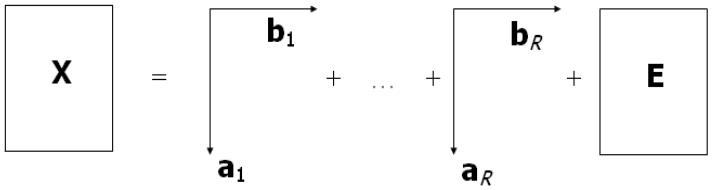


Figure 2.1: Two-way decomposition

To find an optimal solution of (2.1), there is a commonly employed method that is PCA by means of SVD.

2.1.1 Singular Value Decomposition (SVD)

The SVD of a matrix is defined as follows. Let $\mathbf{X}$ be an arbitrary $I \times J$ matrix, $I \geq J$, of rank R, $R \leq J$. Then the SVD of $\mathbf{X}$ is of the form as $\mathbf{X} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, where $\mathbf{U}$ is a $I \times J$ matrix with $\mathbf{U}^T\mathbf{U} = \mathbf{I}_J$, $\mathbf{V}$ is a $J \times J$ matrix with $\mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = \mathbf{I}_J$, and $\mathbf{S} = \text{diag}\{s_1, ..., s_J\}$ with $s_1 \geq \cdots \geq s_J \geq 0$. The elements $s_1, \ldots, s_J$ are called the singular values of $\mathbf{X}$. It can be seen that SVD has the following interesting property.

Lemma 2.1.1 *The rank of a matrix equals the number of its non-zero singular values.*

The SVD of an arbitrary matrix $\mathbf{X}$ always exists. In the case in which $\mathbf{X}$ is not of full column rank, i.e. $R < J$, a parsimonious version of the SVD of $\mathbf{X}$

is available by the simplified expression $\mathbf{X} = \mathbf{U}_R\mathbf{S}_R\mathbf{V}_R^T$, where $\mathbf{U}_R$ (respectively $\mathbf{V}_R$) is formed by taking the first R columns of $\mathbf{U}$ (respectively $\mathbf{V}$), and $\mathbf{S}_R$ is the upper left $R \times R$ submatrix of $\mathbf{S}$.

To find the best low-rank approximation of a matrix, Eckart and Young (1936) provided a method using the SVD. They showed that every matrix always has the best low-rank approximation as the truncated SVD of this matrix. This result is as follows.

Theorem 2.1.2 (Eckart & Young (1936)) *Let* $\mathbf{Y}$ *be given by the truncated SVD of* $\mathbf{X}$*, i.e.* $\mathbf{Y} = \sum_{r=1}^{R} s_r(\mathbf{u}_r \circ \mathbf{v}_r)$ *for some* $R \leq \operatorname{rank}(\mathbf{X})$*. Then* $\mathbf{Y}$ *is a best rank-*R *approximation of* $\mathbf{X}$*, that means* $||\mathbf{X} - \mathbf{Y}||^2 = \sum_{i,j}(x_{ij} - y_{ij})^2$ *is minimal over* $\mathbf{Y}$ *of* $\operatorname{rank} R$*.*

For $\mathbf{X} = \mathbf{Y}$ it can be seen that the rank R is indeed the smallest number of rank-1 matrices whose sum equals $\mathbf{X}$.

2.1.2 Principal Component Analysis (PCA)

Principal component analysis (PCA) is a popular statistical method which was invented in 1901 by Pearson (1901). Now, it is mostly used as a tool in exploratory data analysis and for making predictive models. It can convert a set of observations of possibly correlated variables into a reduced set of values of uncorrelated variables, called principal components, that capture most of the information contained in the original data. This allows to reduce the dimension of the original data set while losing the smallest possible amount of information.

Formally, let $\mathbf{X}$ be a $I \times J$ data matrix with standardized scores (each column has mean 0 and variance 1) of I individuals on J variables. The PCA of $\mathbf{X}$ with R components is to find a $I \times R$ orthogonal matrix $\mathbf{A}$ and a $J \times R$ matrix $\mathbf{B}$ such that $\mathbf{X}$ can be expressed as $\mathbf{X} = \mathbf{A}\mathbf{B}^T + \mathbf{E}$, where $||\mathbf{E}||^2$ is minimal. The

matrix $\mathbf{A}$ is called the component score matrix, and its columns are the principal components. The matrix $\mathbf{B}$ is called the loading matrix, the loadings are the weights that allow to reconstruct the original variables as linear combinations of the principal components. The pair $(\mathbf{A}, \mathbf{B})$ is called a PCA solution.

Also, PCA is presented in an other way by the following expression

$$\mathbf{X} = \sum_{r=1}^{R} \mathbf{a}_r \mathbf{b}_r^T + \mathbf{E},$$

which shows that PCA tries to fit $\mathbf{X}$ as a sum of R rank-one matrices.

The objective of PCA is minimizing $||\mathbf{E}||^2 = ||\mathbf{X} - \mathbf{A}\mathbf{B}^T||^2$. Since $\text{rank}(\mathbf{A}\mathbf{B}^T) \leq R$, it follows that $\mathbf{A}\mathbf{B}^T$ is the truncated SVD of $\mathbf{X}$, that is if $\mathbf{U}_R\mathbf{S}_R(\mathbf{V}_R)^T$ is the truncated SVD of $\mathbf{X}$ then $\mathbf{A} = I^{\frac{1}{2}}\mathbf{U}_R$ and $\mathbf{B}^T = I^{-\frac{1}{2}}\mathbf{S}_R(\mathbf{V}_R)^T$. The minimum number of components that give perfect fit is the number of non-zero singular values of $\mathbf{X}$, that is $\text{rank}(\mathbf{X})$. Therefore, it is not necessary to take a number of components R larger than the number of variables J. In fact, R is usually taken much smaller than J.

Due to the decreasing ordering of singular values of $\mathbf{X}$ in SVD, the first principal component explains the most variance possible, and each succeeding component in turn explains the most variance under the constraint that it be orthogonal to (uncorrelated with) the preceding components. Then the total of explained variances is $\text{tr}(\mathbf{B}\mathbf{B}^T) = \text{tr}(I^{-1}\mathbf{S}_R^2)$, with $\text{tr}(\mathbf{S}_R^2)$ being the sum of squares of the R largest singular values of $\mathbf{X}$.

A PCA solution is not unique, but $\mathbf{A}$ and $\mathbf{B}$ can be rotated such that $\mathbf{A}\mathbf{B}^T$ remains the same. In PCA, the variables typically have both positive and negative loadings on the same component, which makes for complicated interpretations. By rotating, the principal components are traded for rotated components with easier interpretations. It is well-known that a change of basis for a predictor space leaves the prediction unaffected. Therefore, when the principal compo-

nents are rotated to another set of vectors spanning the same subspace, the matrix $\mathbf{X}$ can be reconstructed exactly as good or bad as before, and nothing changes in the amount of variance of $\mathbf{X}$ that is accounted for by the components. A rotation is either orthogonal or oblique. Both rotations produce new components whose interpretations are better and do not involve any loss of information. Although the orthogonal rotation has the advantage of simplicity, it can be artificial because capacities and properties of individuals are seldom uncorrelated. Meanwhile, the oblique rotation is useful in the most common cases. However, unlike the orthogonal rotation, the oblique rotation does not produce a simple structure, but rather a simple pattern.

Formally, an orthogonal rotation is multiplication by an orthonormal matrix $\mathbf{Q}$ $(R \times R)$, and the rotated model part $(\mathbf{AQ})(\mathbf{BQ})^T$ is identical to $\mathbf{AB}^T$. For an oblique rotation, we have $(\mathbf{AQ})(\mathbf{B}(\mathbf{Q}^{-1})^T)^T$ with $\mathbf{Q}$ $(R \times R)$ non-singular and having length-1 columns. For an overview of rotation criteria and methods, see Browne (2001).

2.2 Three-way decomposition

As a natural generalization of the two-way decompositions, three-way decomposition is studied. However, conditions for existence and uniqueness of the optimal solutions of three-way models are not similar to two-way models. In fact, finding an optimal solution of a three-way model is more difficult.

In 1927, Tensor decompositions originated by Hitchcock (see Hitchcock 1927a, 1927b), and the idea of a multi-way model is attributed to Cattell (1944). These concepts received very little attention until 1966, when Tucker (1966) introduced a form of 3-way PCA in psychometrics, called Turker3 model.

A tensor is a multidimensional array. More formally, an N-way array or

order-N tensor is an element of the tensor product of N vector spaces, each of which has its own coordinate system. Thus, a vector is a 1-way array or an order-1 tensor, a matrix is a 2-way array or an order-2 tensor, and so on.

2.2.1 Rank of a three-way array

There is a similarity between the definition of two-way array rank and of three-way array rank. The rank of a three-way array is also defined as the minimal number of rank-1 arrays whose sum equals the three-way array. A rank-1 array $\underline{\mathbf{X}}$ is the outer product of three non-zero vectors $\mathbf{a} = [a_1, \ldots, a_I]^T$, $\mathbf{b} = [b_1, \ldots, b_J]^T$, $\mathbf{c} = [c_1, \ldots, c_K]^T$ whose entries are given by $x_{ijk} = a_i b_j c_k$, where the subscript ijk corresponds to row i, column j and frontal slice k. This is denoted as $\underline{\mathbf{X}} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$. All slices of $\underline{\mathbf{X}}$ are proportional in each of the three directions. Determining the rank of a three-way array is more complicated than for a two-way array.

Kruskal (1989) provided some very particular features of the rank of three-way arrays. Also, Kruskal (1989) referred to some differences between rank of a two-way array and of a three-way array as: While the two-way array rank is easily determined by straight-forward algorithms, there is no known algorithm to compute the rank of a three-way array and determining the rank of a three-way array is extremely difficult; Also, the maximal rank of two-way arrays of order $I \times J$ is equal to the minimum of I and J while the maximal rank of three-way arrays of order $I \times J \times K$ is generally unknown, very difficult to determine and may be larger than I, J and K; And the rank of a two-way array does not depend on the base field ($\mathbb{R}$ or $\mathbb{C}$) being used but the rank of a three-way array does.

For a multi-way array, a mode-i vector is a vector obtained from the array

by varying the i^{th} index and keeping the other indices fixed. The mode-i rank, denoted rank_i, is defined as the rank of the set of mode-i vectors. For a matrix $\mathbf{X}$, $\text{rank}_1(\mathbf{X})$ is the column-rank, and $\text{rank}_2(\mathbf{X})$ is the row-rank. For the three-way array $\underline{\mathbf{X}}$ of order $I \times J \times K$,

$$\begin{aligned}
\text{rank}_1(\underline{\mathbf{X}}) &= \text{rank}\{\mathbf{x}_{jk}^{(1)} = [x_{1jk} \dots x_{Ijk}]^T, j = 1, \dots, J, k = 1, \dots, K\};\\
\text{rank}_2(\underline{\mathbf{X}}) &= \text{rank}\{\mathbf{x}_{ik}^{(2)} = [x_{i1k} \dots x_{iJk}]^T, i = 1, \dots, I, k = 1, \dots, K\};\\
\text{rank}_3(\underline{\mathbf{X}}) &= \text{rank}\{\mathbf{x}_{ij}^{(3)} = [x_{ij1} \dots x_{ijK}]^T, i = 1, \dots, I, j = 1, \dots, J\}.
\end{aligned}$$

The triplet of these three ranks together is known as the multilinear rank of $\underline{\mathbf{X}}$. We also define three types of slices of $\underline{\mathbf{X}}$ as follows. Mode-12 slices, denoted $\mathbf{X}_k^{(12)}$, $k = 1, \dots, K$, are the matrices obtained from $\underline{\mathbf{X}}$ by varying the first two indices and keeping the last index fixed. Mode-12 slices are also called frontal slices and have the following form

$$\mathbf{X}_k^{(12)} = \begin{pmatrix} x_{11k} & x_{12k} & \cdots & x_{1Jk} \\ x_{21k} & x_{22k} & \cdots & x_{2Jk} \\ \vdots & \vdots & \ddots & \vdots \\ x_{I1k} & x_{I2k} & \cdots & x_{IJk} \end{pmatrix}, \quad k = 1, \dots, K.$$

Mode-23 slices, denoted $\mathbf{X}_i^{(23)}$, $i = 1 \dots I$, are the matrices obtained from $\underline{\mathbf{X}}$ by varying the last two indices and keeping the first index fixed. Mode-23 slices are also called horizontal slices and have the following form

$$\mathbf{X}_i^{(23)} = \begin{pmatrix} x_{i11} & x_{i12} & \cdots & x_{i1K} \\ x_{i21} & x_{i22} & \cdots & x_{i2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{iJ1} & x_{iJ2} & \cdots & x_{iJK} \end{pmatrix}, \quad i = 1, \dots, I.$$

Mode-13 slices, denoted $\mathbf{X}_j^{(13)}$, $j = 1, \dots, J$, are the matrices obtained from $\underline{\mathbf{X}}$ by varying the first index and the third index, and keeping the second index

fixed. Mode-13 slices are also called lateral slices and have the following form

$$\mathbf{X}_j^{(13)} = \begin{pmatrix} x_{1j1} & x_{2j1} & \cdots & x_{Ij1} \\ x_{1j2} & x_{2j2} & \cdots & x_{Ij2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1jK} & x_{2jK} & \cdots & x_{IjK} \end{pmatrix}, \quad j = 1, \ldots, J.$$

For a two-way array $\mathbf{X}$, we have $\text{rank}_1(\mathbf{X}) = \text{rank}_2(\mathbf{X}) = \text{rank}(\mathbf{X})$. However, this does not hold for three-way arrays. For example, let $\underline{\mathbf{X}}$ be a $2 \times 2 \times 2$ array with

$$\mathbf{X}_1^{(12)} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{X}_2^{(12)} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Clearly, $\text{rank}_1(\underline{\mathbf{X}}) = \text{rank}_2(\underline{\mathbf{X}}) = \text{rank}_3(\underline{\mathbf{X}}) = 2$, but $\text{rank}(\underline{\mathbf{X}}) = 3$. In general, $\text{rank}_i(\underline{\mathbf{X}})$ may not be the same for $i = 1, 2, 3$.

The typical rank of a matrix (respectively, array) with given order is the rank that occurs with positive probability, when the elements of the matrix (respectively, array) are sampled from a continuous distribution. For an $I \times J$ matrix, the typical rank equals $\min(I, J)$. For a three-way array, the typical rank may have more than one value; see e.g. Ten Berge (2011).

2.2.2 Candecomp/Parafac model (CP)

Carroll and Chang (1970) and Harshman (1970) independently introduced a form of three-way PCA and named it Candecomp (canonical decomposition) and Parafac (parallel factor analysis), respectively (abbreviated as CP model). A CP decomposition of a three-way array is a decomposition of this array into a sum of rank-1 arrays. The CP model is a sum of a CP decomposition and a residual array. Formally, let $\underline{\mathbf{X}}$ be an $I \times J \times K$ array. Then the CP model of

$\underline{\mathbf{X}}$ can be written as follows.

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} g_r(\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r) + \underline{\mathbf{E}} = \sum_{r=1}^{R} \underline{\mathbf{Y}}^{(r)} + \underline{\mathbf{E}}, \tag{2.2}$$

where $\underline{\mathbf{E}}$ is the residual array, R is the pre-specified number, $\mathbf{a}_r, \mathbf{b}_r, \mathbf{c}_r$ are vectors of length 1, g_r is the weight of the r^{th} component, $\underline{\mathbf{Y}}^{(r)} = g_r(\mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r)$ are rank-1 arrays, called components. For fixed R, the CP decomposition (2.2) is found by minimizing $||\underline{\mathbf{E}}||^2 = \sum\limits_{i,j,k} e_{ijk}^2$. The CP model (2.2) can be graphically depicted as in Figure 2.2.

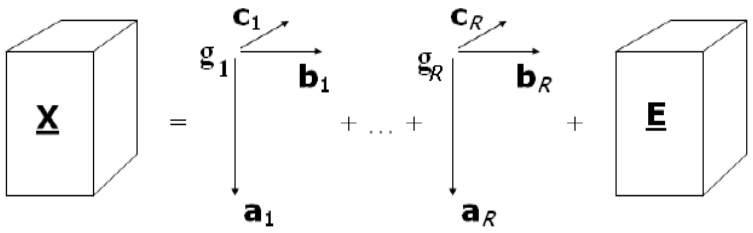


Figure 2.2: Three-way decomposition

We denote $\mathbf{A} = [\mathbf{a}_1|...|\mathbf{a}_R]$ $(I \times R)$, $\mathbf{B} = [\mathbf{b}_1|...|\mathbf{b}_R]$ $(J \times R)$ and $\mathbf{C} = [\mathbf{c}_1|..|\mathbf{c}_R]$ $(K \times R)$, called component matrices. Then we can state a matrix notation of the CP model as follows.

$$\mathbf{X}_k^{(12)} = \mathbf{A}\mathbf{C}_k\mathbf{B}^T + \mathbf{E}_k^{(12)}, \quad k = 1, \ldots, K, \tag{2.3}$$

$$\mathbf{X}_i^{(23)} = \mathbf{B}\mathbf{A}_i\mathbf{C}^T + \mathbf{E}_i^{(23)}, \quad i = 1, \ldots, I, \tag{2.4}$$

$$\mathbf{X}_j^{(13)} = \mathbf{C}\mathbf{B}_j\mathbf{A}^T + \mathbf{E}_j^{(13)}, \quad j = 1, \ldots, J, \tag{2.5}$$

where $\mathbf{C}_k$ $(R \times R)$ is a diagonal matrix with the k^{th} elements of the vectors $g_r\mathbf{c}_r$ on its diagonal, $\mathbf{A}_i$ $(R \times R)$ is a diagonal matrix with the i^{th} elements of the vectors $g_r\mathbf{a}_r$ on its diagonal, $\mathbf{B}_j$ $(R \times R)$ is a diagonal matrix with the j^{th} elements of the vectors $g_r\mathbf{b}_r$ on its diagonal, $r = 1, \ldots, R$. Hence, without loss

of generality, it can be assumed that the weights g_r are absorbed in $\mathbf{A}$ or $\mathbf{B}$ or $\mathbf{C}$.

There is also another way to write the CP model. That is a matrix unfolding notation by using Khatri-Rao product as in the following three forms

$$\begin{aligned}\mathbf{X}_{(JK\times I)} &= \left(\mathbf{X}_1^{(12)}\cdots\mathbf{X}_K^{(12)}\right)^T = \begin{pmatrix}\mathbf{B}\mathbf{C}_1\\ \vdots \\ \mathbf{B}\mathbf{C}_K\end{pmatrix}\mathbf{A}^T + \mathbf{E}_{(JK\times I)}\\ &= (\mathbf{C}\odot\mathbf{B})\mathbf{A}^T + \mathbf{E}_{(JK\times I)}, \end{aligned} \tag{2.6}$$

where g_r is absorbed in $\mathbf{C}$;

$$\begin{aligned}\mathbf{X}_{(IK\times J)} &= \left(\mathbf{X}_1^{(23)}\cdots\mathbf{X}_I^{(23)}\right)^T = \begin{pmatrix}\mathbf{C}\mathbf{A}_1\\ \vdots \\ \mathbf{C}\mathbf{A}_I\end{pmatrix}\mathbf{B}^T + \mathbf{E}_{(IK\times J)}\\ &= (\mathbf{A}\odot\mathbf{C})\mathbf{B}^T + \mathbf{E}_{(IK\times J)}, \end{aligned} \tag{2.7}$$

where g_r is absorbed in $\mathbf{A}$;

$$\begin{aligned}\mathbf{X}_{(IJ\times K)} &= \left(\mathbf{X}_1^{(13)}\cdots\mathbf{X}_J^{(13)}\right)^T = \begin{pmatrix}\mathbf{A}\mathbf{B}_1\\ \vdots \\ \mathbf{A}\mathbf{B}_J\end{pmatrix}\mathbf{C}^T + \mathbf{E}_{(IJ\times K)}\\ &= (\mathbf{B}\odot\mathbf{A})\mathbf{C}^T + \mathbf{E}_{(IK\times J)}, \end{aligned} \tag{2.8}$$

where g_r is absorbed in $\mathbf{B}$. The notation $\odot$ stands for Khatri-Rao product and is defined as column-wise Kronecker product of matrices, i.e.

$$\mathbf{A}\odot\mathbf{B} = [\mathbf{a}_1\otimes\mathbf{b}_1|\mathbf{a}_2\otimes\mathbf{b}_2|\cdots|\mathbf{a}_R\otimes\mathbf{b}_R].$$

Clearly, there is a similarity between model (2.1) and model(2.2), also between Figure 2.1 and Figure 2.2. A triplet $(\mathbf{A},\mathbf{B},\mathbf{C})$ that satisfies the equation

(2.2) is called a CP solution or a rank-R approximation of $\underline{\mathbf{X}}$. If $R = \text{rank}(\underline{\mathbf{X}})$ then $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is a perfect fitting solution of CP, that is $\underline{\mathbf{E}} = \underline{\mathbf{O}}$. If $R < \text{rank}(\underline{\mathbf{X}})$ and $||\underline{\mathbf{E}}||^2$ is minimal then $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is called an optimal CP solution or a best rank-R approximation of $\underline{\mathbf{X}}$. Thus, the rank of $\underline{\mathbf{X}}$ is the smallest number R for which $\underline{\mathbf{X}}$ has a full CP decomposition.

Contrary to the two-way case, we do not have a three-way SVD that yields a best rank-R approximation to a three-way array $\underline{\mathbf{X}}$ for any R. In order to find the R components and their weights in CP, an iterative algorithm is needed which minimizes $||\underline{\mathbf{E}}||^2$. For an overview and comparison of CP algorithms, see Tomasi and Bro (2006) and Comon, Luciani, and De Almeida (2009).

2.2.3 Tucker3 model

The Tucker3 model has been proposed by Tucker (1966). It is defined as

$$\underline{\mathbf{X}} = \sum_{r=1}^{R} \sum_{p=1}^{P} \sum_{q=1}^{Q} g_{rpq} (\mathbf{a}_r \circ \mathbf{b}_p \circ \mathbf{c}_q) + \underline{\mathbf{E}}. \tag{2.9}$$

It is easy to see that the CP model (2.2) is a special case of the aforementioned Tucker3 model. Indeed, for $R = P = Q$ and $g_{rpq} = 0$ if $(r, p, q) \neq (r, r, r)$, model (2.9) is identical to (2.2). The $R \times P \times Q$ array $\underline{\mathbf{G}}$ with entries g_{rpq} is referred to as the core array. Usually, the component matrices $\mathbf{A}$ $(I \times R)$, $\mathbf{B}$ $(I \times P)$ and $\mathbf{C}$ $(K \times Q)$ are restricted (without loss of fit) to be column-wise orthonormal. The matrix notation of Tucker3 model can be stated as follows.

$$\mathbf{X}_k^{(12)} = \mathbf{A} \left(\sum_{q=1}^{Q} c_{kq} \mathbf{G}_q \right) \mathbf{B}^T + \mathbf{E}_k^{(12)}, \quad k = 1 \ldots K,$$

where $\mathbf{G}_q$ $(R \times P)$ is the q^{th} frontal slice of $\underline{\mathbf{G}}$. We can rewrite the matrix notation of Tucker3 model as follows.

$$\begin{aligned}
\mathbf{X}_{JK\times I} = \left(\mathbf{X}_1^{(12)} \cdots \mathbf{X}_K^{(12)}\right)^T &= \begin{pmatrix} \sum_{q=1}^{Q} c_{1q}\mathbf{B}\mathbf{G}_q^T \\ \vdots \\ \sum_{q=1}^{Q} c_{Kq}\mathbf{B}\mathbf{G}_q^T \end{pmatrix} \mathbf{A}^T + \mathbf{E}_{(JK\times I)} \\
&= \begin{pmatrix} c_{11}\mathbf{B} & \cdots & c_{1Q}\mathbf{B} \\ \vdots & & \vdots \\ c_{K1}\mathbf{B} & \cdots & c_{KQ}\mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{G}_1^T \\ \vdots \\ \mathbf{G}_Q^T \end{pmatrix} \mathbf{A}^T + \mathbf{E}_{(JK\times I)} \\
&= (\mathbf{C}\otimes\mathbf{B})\mathbf{G}^T\mathbf{A}^T + \mathbf{E}_{(JK\times I)}, \qquad (2.10)
\end{aligned}$$

where $\mathbf{G} = [\mathbf{G}_1 \cdots \mathbf{G}_Q]$. Notation of Tucker3 model is also stated as $\underline{\mathbf{X}} = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \underline{\mathbf{G}} + \underline{\mathbf{E}}$, which is defined as follows.

$$\underline{\mathbf{X}} = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \underline{\mathbf{G}} \Leftrightarrow x_{ijk} = \sum_{r=1}^{R}\sum_{p=1}^{P}\sum_{q=1}^{Q} a_{ir}\ b_{jp}\ c_{kq}\ g_{rpq}.$$

If $P = Q = R$ and $\underline{\mathbf{G}}$ is a super-diagonal $R \times R \times R$ array with following R frontal slices

$$\mathbf{G}_1 = \begin{pmatrix} g_1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \ \ldots, \ \mathbf{G}_R = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g_R \end{pmatrix},$$

then the notation of CP model (2.2) is stated as $\underline{\mathbf{X}} = (\mathbf{A}, \mathbf{B}, \mathbf{C}) \cdot \underline{\mathbf{G}} + \underline{\mathbf{E}}$. The Tucker3 model can be fitted by alternating least squares (Kroonenberg & De Leeuw, 1980), see section 2.2.5. For $R \leqslant I$, $P \leqslant J$, $Q \leqslant K$, a best-fitting Tucker3 model is a best multilinear rank-(R, P, Q) approximation of $\underline{\mathbf{X}}$. Algorithms can be found in Savas and Lim (2010) and Ishteva, Absil, Van Huffel,

and De Lathauwer (2011). Under additional constraints, the Tucker3 model can be viewed as a multilinear SVD (De Lathauwer, De Moor, & Vandewalle, 2000). For guidelines on fitting the Tucker3 model see Kiers and Van Mechelen (2001).

2.2.4 Computation of CP via Alternating Least Squares (ALS)

In this section, we can assume that the weights in the model (2.2) are absorbed in the matrix $\mathbf{C}$. The simplest algorithm for fitting the CP model is called alternating least squares and works as follows.

1. (Random) starting values for $(\mathbf{A}, \mathbf{B}, \mathbf{C})$, we have

$$\varepsilon_0 = l(\mathbf{A}, \mathbf{B}, \mathbf{C}) = \sum_{k=1}^{K} ||\mathbf{X}_k^{(12)} - \mathbf{A}\mathbf{C}_k\mathbf{B}^T||^2.$$

2. Find the best $\mathbf{A}$ for fixed $\mathbf{B}$ and $\mathbf{C}$ (regression)

$$\mathbf{A}^T = \left(\sum_{k=1}^{K} \mathbf{C}_k\mathbf{B}^T\mathbf{B}\mathbf{C}_k\right)^{-1} \left(\sum_{k=1}^{K} \mathbf{C}_k\mathbf{B}^T \left(\mathbf{X}_k^{(12)}\right)^T\right).$$

3. Find the best $\mathbf{B}$ for fixed $\mathbf{A}$ and $\mathbf{C}$ (regression)

$$\mathbf{B}^T = \left(\sum_{k=1}^{K} \mathbf{C}_k\mathbf{A}^T\mathbf{A}\mathbf{C}_k\right)^{-1} \left(\sum_{k=1}^{K} \mathbf{C}_k\mathbf{A}^T \left(\mathbf{X}_k^{(12)}\right)^T\right).$$

4. Find the best $\mathbf{C}$ for fixed $\mathbf{A}$ and $\mathbf{B}$ (regression)

$$\mathbf{c}_k = \left((\mathbf{A} \odot \mathbf{B})^T(\mathbf{A} \odot \mathbf{B})\right)^{-1} (\mathbf{A} \odot \mathbf{B})^T \mathrm{Vec}(\mathbf{X}_k^{(12)}), \quad k = 1, \ldots, K,$$

   where $\mathbf{c}_k$ is the k^{th} column of $\mathbf{C}$.

5. Stop if relative increase in fit is below some ε. Else repeat steps 2, 3, 4.

2.2.5 Computation of Tucker3 via Alternating Least Squares (ALS)

Set $\mathbf{X} = \mathbf{X}_{I\times JK} = \mathbf{X}^T_{JK\times I}$, where $\mathbf{X}_{JK\times I}$ is determined as in (2.10). Due to Kroonenberg and De Leeuw (1980), the Tucker3 model (2.10) is fitted by minimizing the following function

$$\min_{\mathbf{A},\mathbf{B},\mathbf{C},\mathbf{G}} ||\mathbf{X} - \mathbf{A}\mathbf{G}(\mathbf{C}\otimes\mathbf{B})^T||^2 \tag{2.11}$$

subject to $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ being orthogonal without loss of generality. The ALS algorithm for fitting the Tucker3 model (2.10) works as follows.

1. Initialize $\mathbf{B}$ and $\mathbf{C}$,
2. $\mathbf{A}$ equals first R left singular vectors of $\mathbf{X}_{I\times JK}(\mathbf{C}\otimes\mathbf{B})$ (orthogonal regression)
3. $\mathbf{B}$ equals first P left singular vectors of $\mathbf{X}_{J\times IK}(\mathbf{C}\otimes\mathbf{A})$ (orthogonal regression)
4. $\mathbf{C}$ equals first Q left singular vectors of $\mathbf{X}_{K\times IJ}(\mathbf{B}\otimes\mathbf{A})$ (orthogonal regression)
5. $\mathbf{G} = \mathbf{A}^T\mathbf{X}(\mathbf{C}\otimes\mathbf{B})$,
6. Stop if relative increase in fit is below some ε. Else, repeat steps 2, 3, 4, 5.

2.2.6 Uniqueness properties of CP and Tucker3

One of the most attractive features of CP is its uniqueness property. The uniqueness of a CP solution is usually studied for a given fitted model array $\hat{\underline{\mathbf{X}}} = \underline{\mathbf{X}} - \underline{\mathbf{E}}$. We assume the weights g_r in (2.2) are absorbed in the component matrices. It can be seen that the component matrices $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ in (2.2) can only be unique up

to rescaling and jointly permuting columns of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$. Indeed, if $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is a CP solution of (2.2) then so is $(\overline{\mathbf{A}}, \overline{\mathbf{B}}, \overline{\mathbf{C}}) = (\mathbf{APT}_a, \mathbf{BPT}_b, \mathbf{CPT}_c)$ for a permutation matrix $\mathbf{P}$ and diagonal matrices $\mathbf{T}_a$, $\mathbf{T}_b$, $\mathbf{T}_c$ with $\mathbf{T}_a\mathbf{T}_b\mathbf{T}_c = \mathbf{I}_R$. If these are the only alternatives possible, then the CP solution $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is called essentially unique. For example, for $R = 3$, and $\mathbf{A} = [\mathbf{a}_1\ \mathbf{a}_2\ \mathbf{a}_3]$, $\mathbf{B} = [\mathbf{b}_1\ \mathbf{b}_2\ \mathbf{b}_3]$, $\mathbf{C} = [\mathbf{c}_1\ \mathbf{c}_2\ \mathbf{c}_3]$. If $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ is a CP solution, then for the matrices $\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$, $\mathbf{T}_a = \text{diag}\{\alpha_1, \alpha_2, \alpha_3\}$, $\mathbf{T}_b = \text{diag}\{\beta_1, \beta_2, \beta_3\}$, $\mathbf{T}_c = \text{diag}\{\gamma_1, \gamma_2, \gamma_3\}$, we have $\overline{\mathbf{A}} = \mathbf{APT}_a = [\alpha_1\mathbf{a}_3\ \alpha_2\mathbf{a}_1\ \alpha_3\mathbf{a}_2] = [\overline{\mathbf{a}}_1\ \overline{\mathbf{a}}_2\ \overline{\mathbf{a}}_3]$, $\overline{\mathbf{B}} = \mathbf{BPT}_b = [\beta_1\mathbf{b}_3\ \beta_2\mathbf{b}_1\ \beta_3\mathbf{b}_2] = [\overline{\mathbf{b}}_1\ \overline{\mathbf{b}}_2\ \overline{\mathbf{b}}_3]$, $\overline{\mathbf{C}} = \mathbf{BPT}_c = [\gamma_1\mathbf{c}_3\ \gamma_2\mathbf{c}_1\ \gamma_3\mathbf{c}_2] = [\overline{\mathbf{c}}_1\ \overline{\mathbf{c}}_2\ \overline{\mathbf{c}}_3]$, and $\sum_{j=1}^{3} \overline{\mathbf{a}}_j \circ \overline{\mathbf{b}}_j \circ \overline{\mathbf{c}}_j = (\alpha_1\mathbf{a}_3) \circ (\beta_1\mathbf{b}_3) \circ (\gamma_1\mathbf{c}_3) + (\alpha_2\mathbf{a}_1) \circ (\beta_2\mathbf{b}_1) \circ (\gamma_2\mathbf{c}_1) + (\alpha_3\mathbf{a}_2) \circ (\beta_3\mathbf{b}_2) \circ (\gamma_3\mathbf{c}_2) = \alpha_1\beta_1\gamma_1(\mathbf{a}_3 \circ \mathbf{b}_3 \circ \mathbf{c}_3) + \alpha_2\beta_2\gamma_2(\mathbf{a}_1 \circ \mathbf{b}_1 \circ \mathbf{c}_1) + \alpha_3\beta_3\gamma_3(\mathbf{a}_2 \circ \mathbf{b}_2 \circ \mathbf{c}_2)$. It follows that $\sum_{j=1}^{3} \overline{\mathbf{a}}_j \circ \overline{\mathbf{b}}_j \circ \overline{\mathbf{c}}_j = \sum_{j=1}^{3} \mathbf{a}_j \circ \mathbf{b}_j \circ \mathbf{c}_j$ if and only if $\alpha_j\beta_j\gamma_j = 1$, $j = 1, 2, 3$. Hence, $(\overline{\mathbf{A}}, \overline{\mathbf{B}}, \overline{\mathbf{C}}) = (\mathbf{A}, \mathbf{B}, \mathbf{C})$ if and only if $\mathbf{T}_a\mathbf{T}_b\mathbf{T}_c = \mathbf{I}_3$. A summary of CP uniqueness conditions is given in section 2.3.

For Tucker3 model, it is well-known that the Tucker3 matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$, and $\mathbf{G}$ are not unique. Indeed, if $\mathbf{S}$ $(R \times R)$, $\mathbf{V}$ $(P \times P)$, and $\mathbf{W}$ $(Q \times Q)$ are nonsigular matrices, then

$$\begin{aligned} \mathbf{AG}(\mathbf{C} \otimes \mathbf{B})^T &= \mathbf{AG}(\mathbf{C}^T \otimes \mathbf{B}^T) \\ &= \mathbf{A}\left(\mathbf{S}^T\right)^{-1}\mathbf{S}^T\mathbf{G}(\mathbf{W} \otimes \mathbf{V})(\mathbf{W}^{-1} \otimes \mathbf{V}^{-1})(\mathbf{C}^T \otimes \mathbf{B}^T) \\ &= \mathbf{A}\left(\mathbf{S}^T\right)^{-1}\mathbf{S}^T\mathbf{G}(\mathbf{W} \otimes \mathbf{V})(\mathbf{W}^{-1}\mathbf{C}^T \otimes \mathbf{V}^{-1}\mathbf{B}^T). \end{aligned}$$

This means that we can rotate the component matrices and the core array from $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{G}$ to $\mathbf{A}\left(\mathbf{S}^T\right)^{-1}$, $\mathbf{B}\left(\mathbf{V}^T\right)^{-1}$, $\mathbf{C}\left(\mathbf{W}^T\right)^{-1}$, and $\mathbf{S}^T\mathbf{G}(\mathbf{W} \otimes \mathbf{V})$, respectively. An oblique rotation method to obtain simple structure in the core array can be found in Kiers (1998b). An orthogonal rotation method to obtain simple

structure in the core array and the component matrices can be found in Kiers (1998a).

2.2.7 Best rank-R approximations may not exist

Unlike two-way models such as PCA, an optimal CP solution may not exist. In such cases, convergence of the CP algorithm is extremely slow and some components of the CP solution become more and more proportional as the CP algorithm runs. In the majority of such cases, exactly two components, say $\underline{\mathbf{Y}}^{(s)}$ and $\underline{\mathbf{Y}}^{(t)}$, of the solution display the following pattern

- In all three component matrices, the columns s and t become almost exactly equal up to a sign change, the product of these sign changes being -1.
- The magnitudes of g_s and g_t in (2.2) become arbitrarily large, while $\underline{\mathbf{Y}}^{(s)} + \underline{\mathbf{Y}}^{(t)}$ remains bounded.

This pattern is called a two-factor degeneracy, see Kruskal, Harshman, and Lundy (1989). It can be graphically depicted as in Figure 2.3.

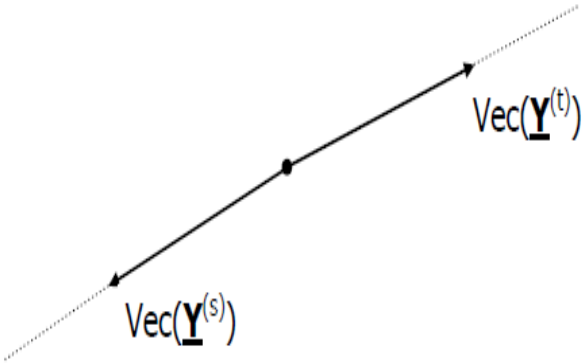


Figure 2.3: Two-factor degeneracy

The contributions of $\underline{\mathbf{Y}}^{(s)}$ and $\underline{\mathbf{Y}}^{(t)}$ diverge in nearly opposite directions. However, their sum $\underline{\mathbf{Y}}^{(s)} + \underline{\mathbf{Y}}^{(t)}$ still contributes to a better fit of the CP decomposition.

Analogous to two-factor degeneracies, also three-factor degeneracies have been encountered, in which the three components $\underline{\mathbf{Y}}^{(s)}$, $\underline{\mathbf{Y}}^{(t)}$ and $\underline{\mathbf{Y}}^{(u)}$ display the following pattern:

- In all three component matrices, the columns s, t and u become almost exactly equal up to a sign change.

- The magnitudes of g_s, g_t and g_u in (2.2) become arbitrarily large, while $\underline{\mathbf{Y}}^{(s)} + \underline{\mathbf{Y}}^{(t)} + \underline{\mathbf{Y}}^{(u)}$ remains bounded.

The sign changes are such that the contribution of two of the factors together nearly cancels the contribution of the third factor, while the sum $\underline{\mathbf{Y}}^{(s)} + \underline{\mathbf{Y}}^{(t)} + \underline{\mathbf{Y}}^{(u)}$ still contributes to a better fit of the CP model. It can be graphically depicted as in Figure 2.4.

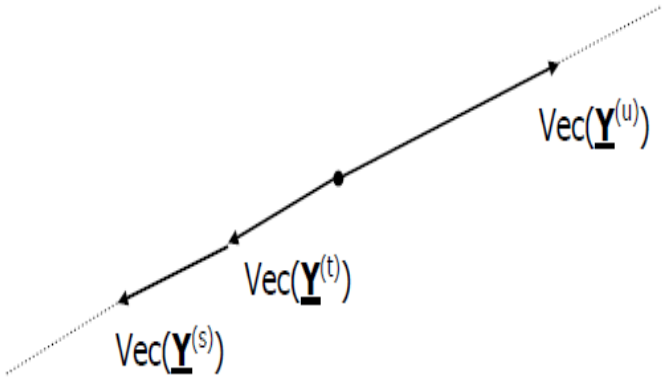


Figure 2.4: Three-factor degeneracy

For example (see Stegeman, 2006), let $\underline{\mathbf{X}}$ be a $3 \times 3 \times 2$ array with two frontal slices as follows

$$\mathbf{X}_1^{(12)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{X}_2^{(12)} = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 1 & 3 \\ 0 & -4 & 2 \end{pmatrix}.$$

For $R = 3$, the following solution is obtained

$$\mathbf{A} = \begin{pmatrix} 0.4864 & 0.4616 & -0.4738 \\ -0.6649 & -0.6535 & 0.6599 \\ -0.5671 & -0.5999 & 0.5835 \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} 0.7236 & -0.6933 & 0.7086 \\ 0.6068 & -0.6507 & 0.6290 \\ 0.3291 & -0.3098 & 0.3203 \end{pmatrix}$$

$$\mathbf{C} = \begin{pmatrix} 0.2893 & -0.2690 & 0.2790 \\ 0.9572 & -0.9631 & 0.9603 \end{pmatrix}$$

and corresponding weights $g_1 = 2258.674$, $g_2 = 2324.58$ and $g_3 = 4579.80$. This is a three-factor degeneracy because $\underline{\mathbf{Y}}^{(1)} \approx \underline{\mathbf{Y}}^{(2)} \approx -\frac{1}{2}\underline{\mathbf{Y}}^{(3)}$.

The degeneracies are also referred to as diverging components or diverging rank-1 terms. Within a CP solution several different groups of diverging components may occur. A more detailed discussion of the non-existence of a best rank-R approximation is provided in section 2.4.

2.3 Uniqueness conditions for CP

In this section, we assume that the weights g_r in (2.2) are absorbed in the component matrices.

Kruskal (1977) has shown that a CP solution $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ of the model (2.2) is essentially unique if

$$k_{\mathbf{A}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2R + 2, \tag{2.12}$$

where $k_{\mathbf{A}}$, $k_{\mathbf{B}}$, $k_{\mathbf{C}}$ denote the Kruskal-ranks of the component matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, respectively. The Kruskal-rank of a matrix is the largest values of m such that every subset of m columns of the matrix is linearly independent. Stegeman and Sidiropoulos (2007) provided a proof of (2.12) that is more accessible than the one by Kruskal (1977). See Rhodes (2010) for an alternative proof.

Ten Berge and Sidiropoulos (2002) showed that Kruskal's uniqueness condition (2.12) is not necessary for CP uniqueness if $R > 3$ (see also Ten Berge & Stegeman, 2006). Weaker uniqueness conditions for CP are given by Jiang and Sidiropoulos (2004) and De Lathauwer (2008). These authors independently examined the case where one of the component matices, say $\mathbf{C}$, has full column rank, i.e. $k_{\mathbf{C}} = R$. Jiang and Sidiropoulos (2004) provided a necessary and sufficient condition of CP uniqueness which unfortunately is not easy to check and also proved that CP uniqueness holds if a matrix $\mathbf{U}_{(\mathbf{A},\mathbf{B})}$ depending on the elements of $\mathbf{A}$ and $\mathbf{B}$ has full column rank. De Lathauwer (2008) showed that, for component matrices $\mathbf{A}$ and $\mathbf{B}$ randomly sampled from a continuous distribution, $\mathbf{U}_{(\mathbf{A},\mathbf{B})}$ has full column rank almost surely if

$$\frac{I(I-1)J(J-1)}{2} \geq R(R-1). \tag{2.13}$$

It can be seen that (2.13) is a weaker condition than (2.12) if $k_{\mathbf{A}} = \min(I, R)$, $k_{\mathbf{B}} = \min(J, R)$ and $k_{\mathbf{C}} = R$. De Lathauwer (2008) also obtained a similar uniqueness condition for the 4-way CP. In Stegeman, Ten Berge, and De Lathauwer (2006), an alternative proof of (2.13) was given using the analysis of Jiang and Sidiropoulos (2004). And Stegeman (2009a) showed that if one of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ has full column rank, then Kruskal's condition (2.12) implies that $\mathbf{U}_{(\mathbf{A},\mathbf{B})}$ has full column rank. Thus, also for non-random $\mathbf{A}$, $\mathbf{B}$ the latter condition is weaker than (2.12).

CP uniqueness conditions are also applied to the CP model of an array with

symmetric slices. For these cases, the Indscal model was introduced by Carroll and Chang (1970) as follows.

$$\mathbf{X}_k^{(12)} = \mathbf{A}\mathbf{C}_k\mathbf{A}^T + \mathbf{E}_k^{(12)}, \quad k = 1, \ldots, K, \tag{2.14}$$

where slices $\mathbf{X}_k^{(12)}$ $(I \times I)$ are symmetric and $\mathbf{C}_k$ $(R \times R)$ are as in (2.3). From Kruskal's uniqueness condition (2.12) for CP, it follows that an Indscal solution is essentially unique if

$$2k_{\mathbf{A}} + k_{\mathbf{C}} \geq 2R + 2 \tag{2.15}$$

For Indscal, the sufficient uniqueness condition of Jiang and Sidiropoulos (2004) is that $\mathbf{U}_{(\mathbf{A},\mathbf{A})}$ has full column rank. The results in Stegeman (2009a) are extended to the Indscal decomposition. Stegeman (2009a) also showed that (2.15) implies that $\mathbf{U}_{(\mathbf{A},\mathbf{A})}$ has full column rank. And Stegeman et al. (2006) partly proved that, for $\mathbf{A}$ randomly sampled from a continuous distribution, $\mathbf{U}_{(\mathbf{A},\mathbf{A})}$ has full column rank almost if

$$\frac{I(I-1)}{2}\left(\frac{I(I-1)}{2}+1\right) - 2\binom{I}{4}1_{\{I\geq 4\}} \geq R(R-1). \tag{2.16}$$

The full proof of (2.16) is given in Stegeman (2011).

Kruskal's conditions (2.12) and (2.15) can be applied to any particular CP and Indscal solution, while conditions (2.13) and (2.16) assume randomly sampled solutions. The question is whether a CP and Indscal solution obtained from an algorithm can be regarded as randomly sampled. Solutions with linearly dependent columns in the component matrices are no exception. This not only happens in the case of "degenerate" solutions (which also occur in Indscal, see Stegeman, 2007), but also in other cases. Hence, the uniqueness conditions (2.13) and (2.16) should be applied with care.

Using the analysis of Jiang and Sidiropoulos (2004), Stegeman (2009a) has obtained improved Kruskal-type uniqueness conditions for CP and Indscal when

$\mathbf{C}$ has full column rank. In particular, if $\text{rank}(\mathbf{A}) = k_{\mathbf{A}}$ or $\text{rank}(\mathbf{B}) = k_{\mathbf{B}}$ or both, then

$$\text{rank}(\mathbf{A}) + \text{rank}(\mathbf{B}) \geq R + 2 \tag{2.17}$$

is sufficient for CP uniqueness. If $k_{\mathbf{A}} < \text{rank}(\mathbf{A})$ and $k_{\mathbf{B}} < \text{rank}(\mathbf{B})$, then the CP uniqueness condition is

$$\max(\text{rank}(\mathbf{A}) + k_{\mathbf{B}}, k_{\mathbf{A}} + \text{rank}(\mathbf{B}) \geq R + 2. \tag{2.18}$$

Conditions (2.17) and (2.18) hold for a particular CP solution and are weaker than Kruskal's condition (2.12) if $k_{\mathbf{C}} = R$. However, (2.17) and (2.18) are stronger than $\mathbf{U}_{(\mathbf{A},\mathbf{B})}$ having full column rank.

For $k_{\mathbf{A}} < \text{rank}(\mathbf{A})$ in Indscal, the improved uniqueness condition (2.18) becomes

$$\text{rank}(\mathbf{A}) + k_{\mathbf{A}} \geq R + 2. \tag{2.19}$$

Condition (2.19) is weaker than Kruskal's condition (2.15) but stronger than $\mathbf{U}_{(\mathbf{A},\mathbf{A})}$ having full column rank.

Recently, Domanov and De Lathauwer (2013a, 2013b) provided a unified treatment of all uniqueness results for three-way CP.

In the Tucker3 model (2.9), the array $\underline{\mathbf{X}}$ is also the sum of a number of rank-1 arrays and a residual array. However, interaction terms between vectors of different modes are taken into account. Their weights are given by the off-diagonal elements of the core array $\underline{\mathbf{G}}$. Contrary to CP, an unrestricted Tucker solution is not unique.

There exists a number of variants of CP and Tucker3 models. Recently, a more general class of multi-way decompositions, called "block decompositions", was introduced by De Lathauwer (2008). In a block decomposition, a multi-way array $\underline{\mathbf{X}}$ is decomposed into R arrays which may have rank larger than

1. De Lathauwer (2008) derived uniqueness conditions of the type (2.12) for a particular solution of the block decomposition.

2.4 Non-existence of a best rank-R approximation

We denote $S_R(I, J, K)$ as the set of all $I \times J \times K$ arrays with rank at most R. Then, a best rank-R approximation of $\underline{\mathbf{X}}$ is defined by solving the following optimization problem.

$$\text{Minimize } ||\underline{\mathbf{X}} - \underline{\mathbf{Y}}||^2 \text{ subject to } \underline{\mathbf{Y}} \in S_R(I, J, K). \tag{2.20}$$

In case of rank($\underline{\mathbf{X}}$) $> R$, an optimal solution of (2.20) will be a boundary point of the set $S_R(I, J, K)$. However, the set $S_R(I, J, K)$ is not closed for $R \geq 2$ (De Silva & Lim, 2008). Therefore, the problem (2.20) may not have an optimal solution. This means that a best rank-R approximation of $\underline{\mathbf{X}}$ or an optimal CP solution may not exist. Non-existence of an optimal CP solution results in diverging rank-1 arrays when an attempt is made to compute a best rank-R approximation to $\underline{\mathbf{X}}$. This was conjectured by Kruskal et al. (1989) and proven by Krijnen, Dijkstra, and Stegeman (2008).

Since diverging components cannot be interpreted, it may become a serious problem in practical use of CP. In simulation studies involving randomly sampled data $\underline{\mathbf{X}}$, diverging components occur very often; see Stegeman (2006, 2008, 2012). Also, it has been proven that all $2 \times 2 \times 2$ arrays of rank 3 have no optimal CP solution for $R = 2$ (see Kruskal et al., 1989). Stegeman (2008) conjectures that for $p \times p \times 2$ arrays $\underline{\mathbf{X}}$ with $\mathbf{X}_2^{(12)} \left(\mathbf{X}_1^{(12)}\right)^{-1}$ having complex eigenvalues, a best rank-p approximation does not exist, i.e. an optimal CP solution does not exist.

In practice, diverging components due to non-existence of an optimal CP solution are often avoided by imposing constraints in CP. For example, Krijnen

et al. (2008) showed that an optimal CP solution will be guaranteed by imposing orthogonality constraints on (one of) the component matrices, and Lim and Comon (2009) showed that an optimal CP solution exists for non-negative $\underline{\mathbf{X}}$ under the restriction of non-negative component matrices. However, these constraints are not suitable for all applications of CP.

There is also a different approach to deal with diverging components. In this approach, instead of considering problem (2.20), one will consider the following problem.

$$\text{Minimize } ||\underline{\mathbf{X}} - \underline{\mathbf{Y}}||^2 \text{ subject to } \underline{\mathbf{Y}} \in \overline{S}_R(I, J, K), \tag{2.21}$$

where $\overline{S}_R(I, J, K)$ denotes the closure of $S_R(I, J, K)$ (i.e., the union of the set and its boundary). Clearly, problem (2.21) always has an optimal solution. And if problem (2.20) has an optimal solution, then it is also an optimal solution of problem (2.21). If problem (2.20) does not have an optimal solution, then the sequence of CP updates will converge to an optimal solution of problem (2.21). This optimal solution is a boundary point of $S_R(I, J, K)$ with rank larger than R. Therefore, to solve problem (2.21), we need to characterize the boundary points of $S_R(I, J, K)$ and need an algorithm to find an optimal boundary point. For $R = 2$, the boundary points are determined by De Silva and Lim (2008), and Rocci and Giordani (2010) showed that problem (2.21) can be solved by fitting a Tucker3 model with column-wise orthonormal component matrices and a constrained $2 \times 2 \times 2$ core array. The set of arrays satisfying this model with perfect fit is equal to $\overline{S}_2(I, J, K)$. For $K = 2$ and $R \leq \min(I, J)$, the boundary points of $S_R(I, J, 2)$ are characterized by Stegeman (see Stegeman, 2006, 2008, 2010), and Stegeman and De Lathauwer (2009) showed that problem (2.21) can be solved for $I \times J \times 2$ arrays by fitting a Generalized Schur Decomposition (GSD). The set of arrays satisfying the GSD with perfect fit is identical to

$\overline{S}_R(I, J, 2)$.

Stegeman (2012, 2013) presented an approach to avoid diverging components by providing a new constructive method to find an optimal solution of problem (2.21) for general $I \times J \times K$ arrays. In these papers, Stegeman showed that a CP decomposition $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ featuring diverging components can be rewritten as a decomposition in block terms where each block term corresponds to a group of diverging components. Moreover, he also showed that if the diverging components occur in groups of two, three or four, then the limiting boundary point $\underline{\mathbf{Z}}$ can be obtained by fitting an appropriate constrained Tucker3 model to $\underline{\mathbf{X}}$, using the block term decomposition of $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ as initial values. The restriction to groups of maximal four diverging components is not severe since larger groups of diverging components occur much less often. In Stegeman (2014), this approach is applied to a three-way dataset of TV-ratings which yields diverging components for $R = 3$.

The methods in Rocci and Giordani (2010) and in Stegeman and De Lathauwer (2009) allow the fitting of a different model than CP. This avoids the slow convergence of a CP algorithm when diverging components occur. In Stegeman (2012, 2013), first CP is fitted. When diverging components occur, an appropriate constrained Tucker3 model is fitted.

2.5 Common factor analysis

2.5.1 Two-way common factor analysis

Let $\mathbf{X}$ be an $I \times J$ matrix containing I observations of J variables. Entry x_{ij} of $\mathbf{X}$ denotes observation i of variable j. We assume the columns of $\mathbf{X}$ (i.e., the J variables) have mean zero. Next, we suppose that in theory the data can be written as the sum of a common part and a unique part: $\mathbf{X} = \mathbf{X}^{(\mathrm{com})} + \mathbf{E}$.

The common part $\mathbf{X}^{(\text{com})}$ contains the part of each variable that correlates with other variables in the data. The unique part $\mathbf{E}$ contains the part of each variable that does not correlate with other variables. The unique part of a variable may contain measurement errors as well as a reliable part measuring a trait that is uncorrelated with any other variable. Both $\mathbf{X}^{(\text{com})}$ and $\mathbf{E}$ have mean-zero columns. We look for a small number of R factors that best summarizes the common part: $\mathbf{X}^{(\text{com})} \approx \mathbf{A}\mathbf{B}^T$, with the $I \times R$ matrix $\mathbf{A}$ containing the factors as columns, and the $J \times R$ matrix $\mathbf{B}$ containing the loadings of the common parts of the variables on the factors. Hence, entry b_{jr} of $\mathbf{B}$ is the loading on factor r of the common part of variable j. For perfect fit, the $J \times J$ covariance matrix of the above factor analysis model is given by

$$\mathbf{\Sigma} = \mathbf{B}\mathbf{\Phi}\mathbf{B}^T + \mathbf{U}, \tag{2.22}$$

where $\mathbf{\Sigma} = I^{-1}\mathbf{X}^T\mathbf{X}$ is the data covariance matrix , $\mathbf{\Phi} = I^{-1}\mathbf{A}^T\mathbf{A}$ is the factor covariance matrix, and $\mathbf{U} = I^{-1}\mathbf{E}^T\mathbf{E}$ is the diagonal matrix of unique variances. Note that $\mathbf{U}$ is diagonal because the unique part of a variable is not correlated with any other variable. The factors $\mathbf{A}$ are usually scaled such that they have variance 1, which makes $\mathbf{\Phi}$ the factor correlation matrix. If the factors are chosen uncorrelated (also called orthogonal), then $\mathbf{\Phi} = \mathbf{I}_R$. Otherwise, the factors are called oblique. Since the factors correspond to the common part of the data, the model (2.22) is known as the common factor model, e.g. Spearman (1904) and Thurstone (1935). The diagonal entries of $\mathbf{\Sigma} - \mathbf{U}$ are the variances of the common parts of the variables, and are called communalities or common variances. The diagonal entries of $\mathbf{B}\mathbf{\Phi}\mathbf{B}^T$ are called the estimated common variances. The diagonal entries of $\mathbf{U}$ are called the unique variances.

Given the data covariance matrix $\mathbf{\Sigma}$, we need to determine matrices $\mathbf{B}$, $\mathbf{\Phi}$, and $\mathbf{U}$ such that the right-hand side of (2.22) is a best approximation of $\mathbf{\Sigma}$

in some sense. Then the triplet $(\mathbf{B}, \mathbf{\Phi}, \mathbf{U})$ is called the solution of (2.22). The solution of (2.22) can be found by minimizing the sum-of-squares of $\mathbf{\Sigma} - \mathbf{B\Phi B}^T - \mathbf{U}$ (the MINRES method of Harman & Jones, 1966), by using the Maximum Likelihood principle (MLFA of Jöreskog, 1969), or using the Minimum Rank Factor Analysis (MRFA of Ten Berge & Kiers, 1991).

The MRFA algorithm computes the unique variances $\mathbf{U}$ such that $\mathbf{U}$ is nonnegative, $\mathbf{\Sigma} - \mathbf{U}$ is a covariance matrix, and the unexplained common variance in $\mathbf{\Sigma} - \mathbf{U} \approx \mathbf{B\Phi B}^T$ is minimized. Since $\mathbf{\Sigma} - \mathbf{U}$ is a covariance matrix, all its eigenvalues are nonnegative. Then the best approximation $\mathbf{B\Phi B}^T$ is obtained from the R largest eigenvalues and associated eigenvectors of $\mathbf{\Sigma} - \mathbf{U}$, and the minimum unexplained common variance in $\mathbf{\Sigma} - \mathbf{U} \approx \mathbf{B\Phi B}^T$ is equal to the sum of the $J - R$ smallest eigenvalues of $\mathbf{\Sigma} - \mathbf{U}$; see Eckart and Young (1936). The advantage of MRFA is that we have proper communalities and we can compute the percentage of explained common variance as

$$100 \cdot \frac{\text{trace}(\mathbf{B\Phi B}^T)}{\text{trace}(\mathbf{\Sigma} - \mathbf{U})}, \tag{2.23}$$

where trace$(\cdot)$ is defined as the sum of the diagonal entries of a matrix, which is equal to the sum of the eigenvalues of the matrix. The numerator of (2.23) equals the sum of the estimated common variances. The denominator equals the sum of the communalities (common variances), under the condition that the eigenvalues of $\mathbf{\Sigma} - \mathbf{U}$ are nonnegative. To sum up, for a fixed number of R factors, MRFA minimizes the amount of common variance left unexplained under the constraint of proper communalities.

In a solution $(\mathbf{B}, \mathbf{\Phi}, \mathbf{U})$ of (2.22), the loadings matrix $\mathbf{B}$ and factor correlation matrix $\mathbf{\Phi}$ are not unique. Indeed, we have

$$\mathbf{B\Phi B}^T = (\mathbf{BQD}^{-1})(\mathbf{DQ}^T\mathbf{\Phi QD})(\mathbf{BQD}^{-1})^T, \tag{2.24}$$

where $\mathbf{Q}$ is an $R \times R$ orthonormal matrix (i.e., $\mathbf{Q}\mathbf{Q}^T = \mathbf{I}_R$), and $\mathbf{D}$ is an $R \times R$ diagonal matrix such that the diagonal of $\mathbf{D}\mathbf{Q}^T\mathbf{\Phi}\mathbf{Q}\mathbf{D}$ contains only ones. Hence, an alternative loadings matrix is $\mathbf{B}\mathbf{Q}\mathbf{D}^{-1}$ and an alternative factor correlation matrix is $\mathbf{D}\mathbf{Q}^T\mathbf{\Phi}\mathbf{Q}\mathbf{D}$. Here, matrix $\mathbf{Q}$ rotates the factors, while $\mathbf{D}$ rescales them. The inverse rotation and rescaling is applied to the loadings matrix. In practice, a rotation is often applied such that factor interpretation via the new $\mathbf{B}\mathbf{\Phi}$ is made easier; see e.g. Browne (2001).

CFA differs from PCA. That is, CFA takes into account the unique part of each variable while this is neglected in PCA. The unique part may be seen as measurement error or as a part of the variable that does not correlate with other variables or as a combination of both. PCA can be favored because of its computational simplicity and manifest component approach, whereas CFA is computationally more difficult and features latent factors. However, PCA and CFA are said to yield similar estimated loadings in simulation studies, although this is not generally true and theoretical results prove only asymptotical similarity (e.g., Ogasawara, 2000) or for the case of equal unique variances (e.g., Tipping & Bishop, 1999).

2.5.2 Three-way common factor analysis

Let $\mathbf{X}_1, \ldots, \mathbf{X}_K$ be frontal slices of a three-way data of size $I \times J \times K$. Here $\mathbf{X}_k$ is a matrix containing I observations of J variables for occasion k or under condition k, for $k = 1, \ldots, K$. Entry x_{ijk} of $\mathbf{X}_k$ denotes observation i of variable j under condition k. We assume the columns of $\mathbf{X}_k$ have mean zero for all k. As in two-mode factor analysis, we write $\mathbf{X}_k$ as the sum of a common part and a unique part: $\mathbf{X}_k = \mathbf{X}_k^{(\text{com})} + \mathbf{E}_k$, for $k = 1, \ldots, K$. Moreover, we look for a small number of R factors that best summarizes the common parts: $\mathbf{X}_k^{(\text{com})} \approx \mathbf{A}\mathbf{B}_k^T$,

where the factors $\mathbf{A}$ are the same for all k, but the loadings $\mathbf{B}_k$ may be different. For all $\mathbf{X}_k$ together, and perfect fit, we have the following factor model:

$$\mathbf{X}_{(I\times JK)} = \mathbf{A}\begin{bmatrix}\mathbf{B}_1\\ \vdots \\ \mathbf{B}_K\end{bmatrix}^T + \mathbf{E}_{(I\times JK)}, \tag{2.25}$$

where $\mathbf{X}_{(I\times JK)} = [\mathbf{X}_1 \cdots \mathbf{X}_K]$ and $\mathbf{E}_{(I\times JK)} = [\mathbf{E}_1 \cdots \mathbf{E}_K]$. Analogous to (2.22), the covariance model corresponding to (2.25) is

$$\mathbf{\Sigma} = \begin{bmatrix}\mathbf{\Sigma}_{11} & \cdots & \mathbf{\Sigma}_{1J}\\ \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{J1} & \cdots & \mathbf{\Sigma}_{JJ}\end{bmatrix} = \begin{bmatrix}\mathbf{B}_1\\ \vdots \\ \mathbf{B}_K\end{bmatrix}\mathbf{\Phi}\begin{bmatrix}\mathbf{B}_1\\ \vdots \\ \mathbf{B}_K\end{bmatrix}^T + \mathbf{U}, \tag{2.26}$$

where $\mathbf{\Sigma} = I^{-1}\mathbf{X}^T_{(I\times JK)}\mathbf{X}_{(I\times JK)}$ is the data covariance matrix, $\mathbf{\Sigma}_{kl} = I^{-1}\mathbf{X}_k^T\mathbf{X}_l$ contains the covariances between the I variables for conditions k and l, the factor covariance matrix $\mathbf{\Phi} = I^{-1}\mathbf{A}^T\mathbf{A}$ is as in (2.22), and $\mathbf{U} = I^{-1}\mathbf{E}^T_{(I\times JK)}\mathbf{E}_{(I\times JK)}$ is the diagonal matrix of unique variances. Note that $\mathbf{\Sigma}$ and $\mathbf{U}$ have size $JK\times JK$, that $\mathbf{\Sigma}_{kl}$ has size $J\times J$, and that $\mathbf{\Sigma}_{kl} = \mathbf{\Sigma}_{lk}^T$.

There is an approach to three-mode factor analysis in the literature that is based on the Tucker3 model (2.10). In this model, each mode of the data has its own components, and their interaction strengths are given by numbers g_{rpq} of the so called core array. For our three-way data, suppose we have R components for the I observations, P components for the J variables, and Q components for the K conditions. The covariance model corresponding to (2.10) is given as

$$\mathbf{\Sigma} \approx (\mathbf{C}\otimes\mathbf{B})\mathbf{G}^T\mathbf{\Phi}\mathbf{G}(\mathbf{C}\otimes\mathbf{B})^T + \mathbf{U}. \tag{2.27}$$

It is well known that the CP model is a special case of the Tucker3 model in which $R = P = Q$ and $g_{rrr} = 1$ and $g_{rpq} = 0$ otherwise. The solution of the

Tucker3 model is not unique. All of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ can be rotated, with inverse transformations applied to the interaction strengths in $\mathbf{G}$, without affecting the model part $\mathbf{AG}(\mathbf{C} \otimes \mathbf{B})^T$. In chapter 5 we consider covariance model (2.27).

We propose a new approach to three-mode factor analysis that is based on the CP model (2.6). The covariance model corresponding to (2.6) is then

$$\mathbf{\Sigma} \approx (\mathbf{C} \odot \mathbf{B})\mathbf{\Phi}(\mathbf{C} \odot \mathbf{B})^T + \mathbf{U}. \tag{2.28}$$

Our approach will be presented in detail in chapter 4.

2.6 Multi-set component and factor analysis

2.6.1 Parafac2

In this section, we consider a set of data matrices in which each data matrix has the same column units, but different row units. Such data set is called multi-set or multi-block. In particular, we consider multi-set data in which the same variables are observed for several different populations or subpopulations. Let $\mathbf{X}_k$ $(I_k \times J)$ be the data matrix of the sample from (sub)population k, $k = 1 \dots K$. If $I_1 = \cdots = I_K$, then we obtain a three-way dataset that could be modelled by CP or Tucker3. In the case $I_1, \dots, I_K$ are not the same, CP and Tucker3 cannot be used. Next, we discuss Parafac2, which is an adaptation of CP to multi-set data. The direct model of Parafac2 is given in Kiers, Ten Berge, and Bro (1999) as follows

$$\mathbf{X}_k = \mathbf{A}_k\mathbf{C}_k\mathbf{B}^T + \mathbf{E}_k, \quad k = 1 \dots K, \tag{2.29}$$

where $\mathbf{X}_k$ is an $I_k \times J$ matrix consisting I_k observations on J variables in sample k, $\mathbf{A}_k$ is an $(I_k \times R)$ matrix of component scores such that $I_k^{-1}\mathbf{A}_k^T\mathbf{A}_k = \mathbf{\Phi}$ is constant over k with a positive semi-definite matrix $\mathbf{\Phi}$, $\mathbf{B}$ $(J \times R)$ is loading

matrix, $\mathbf{C}_k$ is an $R \times R$ diagonal matrix containing the weights for sample k, and $\mathbf{E}_k$ is an $I_k \times J$ matrix of residual on sample k.

The model (2.29) can be graphically depicted as in Figure (2.5)

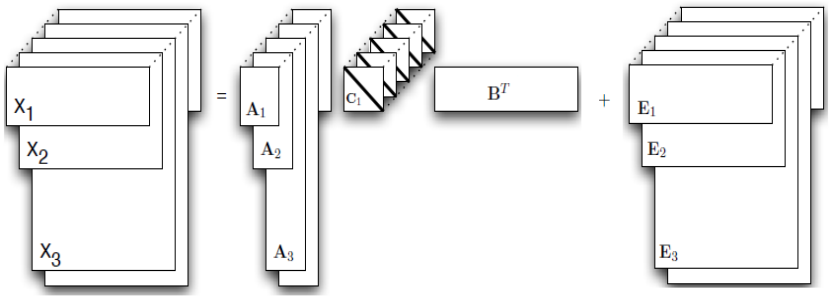


Figure 2.5: The direct Parafac2 model

Harshman (1972) introduced the indirect form of Parafac2 as

$$\mathbf{\Sigma}_k = \mathbf{B}\mathbf{C}_k\mathbf{\Phi}\mathbf{C}_k\mathbf{B}^T + \tilde{\mathbf{E}}_k, \quad k = 1 \dots K, \tag{2.30}$$

where $\mathbf{\Sigma}_k = I_k^{-1}\mathbf{X}_k^T\mathbf{X}_k$ is the covariance matrix, $\mathbf{\Phi} = I_k^{-1}\mathbf{A}_k^T\mathbf{A}_k$ is the factor correlation matrix, and $\tilde{\mathbf{E}}_k$ is the residual in sample k. Note that (2.30) can be obtained from (2.29) when $\mathbf{A}_k^T\mathbf{E}_k = \mathbf{O}$ for $k = 1 \dots K$.

2.6.2 Simultaneous Component Analysis models

Timmerman and Kiers (2003) presented a class of four simultaneous component analysis (SCA) models. Let $\mathbf{X}_k$ be an $(I_k \times J)$ matrix containing scores of sample k on J variables measured at I_k observations, $k = 1 \dots K$. In each of the four SCA models, each sample is decomposed into a few component scores and a loading matrix, where the loading matrix is assumed common for all samples. Difference between the four SCA models is the constraints imposed on the covariances of the component scores. The four SCA models are given as follows.

The model for SCA with invariant Pattern (SCA-P) is given as

$$\mathbf{X}_k = \mathbf{F}_k\mathbf{B}^T + \mathbf{E}_k, \quad k = 1 \ldots K, \tag{2.31}$$

where $\mathbf{F}_k$ $(I_k \times R)$ is the matrix of component scores in sample k, $\mathbf{B}$ is a $J \times R$ matrix of loadings, and $\mathbf{E}_k$ is a $I_k \times J$ matrix of residual. The component score matrix $\mathbf{F}_k$ is unconstrained.

The model for SCA with Parafac2 constraints (SCA-PF2) is a constrained version of SCA-P model. That is, SCA-PF2 is given by (2.31) with constraint $I_k^{-1}\mathbf{F}_k^T\mathbf{F}_k = \mathbf{C}_k\mathbf{\Phi}\mathbf{C}_k$, where $\mathbf{C}_k$ is a $R \times R$ diagonal matrix and $\mathbf{\Phi}$ is a $R \times R$ positive definite matrix having unit diagonal elements. Hence, SCA-PF2 is exactly the direct Parafac2 model (2.29) (ignoring the "between part" of the SCA models). SCA-PF2 is suitable if the variables indicate concepts that are equally correlated for different samples, and if the degree of variability with respect to these concepts varies between samples.

The model for SCA with Indscal constraints (SCA-IND) is a constrained version of both SCA-P and SCA-PF2. That is SCA-IND is given by (2.31) with constraint $\mathbf{\Phi} = \mathbf{I}_R$. SCA-IND is used if the variables indicate several uncorrelated concepts, and if the samples show differences in variability with respect to these concepts.

The model for SCA with Equal average Cross-product constraints (SCA-ECP) is a constrained version of the SCA-P, SCA-PF2, and SCA-IND models. That is, SCA-ECP is given by (2.31) with constraint $I_k^{-1}\mathbf{F}_k^T\mathbf{F}_k = \mathbf{\Phi}$ for some positive definite $\mathbf{\Phi}$. SCA-ECP is used if all samples show equal variability on the components, and the correlations between the components are equal for all samples. Note that, without loss of generality, the components $\mathbf{F}_k$ can be rotated to $I_k^{-1}\mathbf{F}_k^T\mathbf{F}_k = \mathbf{I}_R$ for all samples.

2.6.3 Multi-set Parafac2 common factor model

In this section we introduce a multi-set factor model based on Parafac2. This model and its estimation procedure are the subject of Chapter 6. Let $\mathbf{X}_k$ be the data matrix of a sample from (sub)population k, $k = 1 \ldots K$. We measure the same J variables in each sample, where we have I_k observations in sample k. We assume that the columns of each $\mathbf{X}_k$ are centered. For R underlying factors, our exploratory factor model is given as

$$\mathbf{X}_k = \mathbf{A}_k \mathbf{C}_k \mathbf{B}^T + \mathbf{E}_k, \quad k = 1 \ldots K, \tag{2.32}$$

where $\mathbf{A}_k$ is an $I_k \times R$ matrix of factor scores in sample k such that $I_k^{-1} \mathbf{A}_k^T \mathbf{A}_k = \mathbf{\Phi}$ is constant over k, $\mathbf{B}$ $(J \times R)$ is a loading matrix common to all samples, $\mathbf{C}_k$ $(R \times R)$ is a diagonal matrix containing the factor strengths in sample k, and $\mathbf{E}_k$ $(I_k \times J)$ is the unique part of sample k. The common part of sample k is thus modelled as $\mathbf{A}_k \mathbf{C}_k \mathbf{B}^T$.

The covariance model corresponding to (2.33) is as follows

$$\mathbf{\Sigma}_k \approx \mathbf{B} \mathbf{C}_k \mathbf{\Phi} \mathbf{C}_k \mathbf{B}^T + \mathbf{U}_k, \quad k = 1 \ldots K, \tag{2.33}$$

where $\mathbf{U}_k$ $(J \times J)$ is a diagonal matrix containing the unique variances for sample k. We refer to (2.33) as the multi-set Parafac2 factor model.

Multi-set Parafac2 factor model (2.33) is the factor analogue of the simultaneous component models SCA-PF2 (for oblique components and direct Parafac2 form) and SCA-IND (for orthogonal components and Indscal form) presented in Timmerman and Kiers (2003). The difference between model (2.33) and the component models SCA-PF2 and SCA-IND is the same as the difference between common factor analysis and principal component analysis for one data matrix.

Chapter 3

Multi-way exploratory analysis of Belief in a Just World (BJW) data

3.1 Introduction

In this chapter, we present an exploratory analysis of questionnaire data on Belief in a Just World. The structure of the questionnaire is such that the data can be considered 4-way data with each item scored on a 7-point scale. We analyze mean differences and correlations of the responses, do PCA on matrix unfoldings of the 4-way data, fit Tucker3 to 3-way parts of the dataset, and fit Tucker4 to the complete dataset. As will be seen, the results of these component models are very similar and easy to interpret.

3.1.1 The BJW concept

The BJW concept is originally proposed by Lerner (1980). Belief in a just world enables people to see the world as stable and safe, which benefits the pursuit of long-term goals. However, this belief may also cause people to misinterpret injustice. Several unidimensional BJW measures have been proposed in the literature. Lipkus, Dalbert, and Siegler (1996) make a distinction between justice for the self and justice for others. A new BJW scale introduced by Stroebe, Postmes, Täuber, Stegeman, and John (2014) also distinguishes between several sources of justice. In this chapter we conduct a statistical and exploratory analysis of a 4-way BJW dataset obtained with the BJW scale of Stroebe et al. (2014). The BJW dataset was kindly provided by Katherine Stroebe.

3.1.2 The new BJW questionnaire with 6 "actors"

In the new BJW questionnaire introduced by Stroebe et al. (2014) the source of justice is specified by means of 6 actors. Each actor has 7 items with scores $1, 2, 3, 4, 5, 6, 7$ measured on 2 perspectives. The 6 actors are: Nature, God, Human Institutions, Other people, Yourself, and Chance. As in Lipkus et al. (1996), the 2 perspectives are: for Other people, and for Yourself. The data is obtained from 2 studies: USlab data, and Mturk data.

The 7 items of each actor and perspective are based on items of Lipkus et al. (1996) (see also section 4.3.2). For each item, the actor is denoted by X. For the perspective "for Other people" the items are:

1. People get what they are entitled to have due to X,
2. I feel that peoples efforts are noticed and rewarded by X,
3. I feel people earn the rewards and punishments they receive from X,

4. I feel that when people meet with fortune this is brought on them by X,
5. X will ensure that people get what they deserve,
6. I feel that the rewards and punishments that people get are fairly given by X,
7. Basically X is fair for people.

Analogously, for each actor the 7 items for the perspective "for Yourself" are:

1. I get what I am entitled to have due to X,
2. I feel that my efforts are noticed and rewarded by X,
3. I feel I earn the rewards and punishments I receive from X,
4. I feel that when I meet with fortune this is brought on me by X,
5. X will ensure that I get what I deserve,
6. I feel that the rewards and punishments that I get are fairly given by X,
7. Basically X is fair for me.

3.1.3 Exploratory analysis of 4-way BJW data

For each study, the new BJW questionnaire as above yields 4-way data of the form subjects $\times$ items $\times$ actors $\times$ perspectives.

The exploratory analysis of the BJW dataset in this chapter is organized as follows. In section 3.2, we present descriptive statistics of the BJW dataset for each study. In section 3.3, we describe the results of preliminary analysis concerning t-tests to compare means and correlations among items and among actors. In section 3.4, we conduct PCA with Varimax rotation on the unfolded

data for each perspective separately. In section 3.5, we fit a Tucker3 model for each perspective separately. Finally, in section 3.6, we fit a Tucker4 model for the two perspectives together.

3.2 The BJW dataset

The BJW here is investigated into two studies. The first study is performed on 117 American students from a Northeastern university. The second study is performed on 246 online volunteers of Amazon's MTurk. In study 1, we have a data set with $N = 109$ subjects after deleting subjects with missing data, and this data is called USlab data. In study 2, the MTurk data is obtained with $N = 236$ subjects after deleting subjects with missing data. The means and standard deviations per actor, perspective, and study, for all items together are as in Table 3.1.

In study 1, the means are generally larger for Yourself than for Others, while the converse is true in study 2. Standard deviations are larger for study 2 than for study 1 (80 of 84 items), probably due to less homogenous sample (larger age range, not all with academic background). Standard deviations are larger for Yourself than for Others for study 2 (39 of 42 items) but not for study 1 (24 of 42 items). Standard deviations are largest for God in both perspectives and studies, perhaps due to differences in ideas about "God".

The frequencies of the observed scores per item, for all actors and perspectives together are between 49 and 326 for study 1, and between 161 and 681 for study 2. The frequencies of score 7 for each item are considerably lower than the rest. However, they are not low enough to consider merging answering categories 6 and 7.

Table 3.1: Means and standard deviations per actor, perspective, and study, for all items together.

	Other (study 1)	**Yourself (study 1)**	**Other (study 2)**	**Yourself (study 2)**
Nature	3.20 (1.71)	3.23 (1.72)	3.06 (1.76)	2.92 (1.81)
God	3.17 (2.04)	3.21 (2.08)	3.92 (2.21)	3.76 (2.28)
Institutions	4.12 (1.35)	4.20 (1.36)	4.09 (1.42)	3.89 (1.51)
Other People	4.36 (1.35)	4.48 (1.31)	4.21 (1.37)	4.06 (1.53)
Yourself	4.56 (1.44)	4.62 (1.50)	4.22 (1.58)	4.29 (1.68)
Chance	3.74 (1.53)	3.62 (1.53)	3.68 (1.70)	3.63 (1.76)

3.3 Preliminary analysis: correlations and t-tests

3.3.1 Preliminary analysis of item and actor correlations

Item correlations across all actors, for perspectives for Other (above the diagonal) and for Yourself (below the diagonal) are reported in Table 3.2.
All items are highly correlated (with item 7 slightly lower) in both studies. Correlations among actors across all items, for perspectives for Others (above the diagonal) and for Yourself (below the diagonal) are reported in Table 3.3.

Table 3.2: Item correlations across all actors, for perspectives for Others (above the diagonal) and for Yourself (below the diagonal).

study 1	**item 1**	**item 2**	**item 3**	**item 4**	**item 5**	**item 6**	**item 7**
item 1	1	0.74	0.76	0.74	0.70	0.70	0.57
item 2	0.79	1	0.80	0.75	0.76	0.77	0.60
item 3	0.74	0.76	1	0.77	0.74	0.79	0.60
item 4	0.76	0.73	0.78	1	0.75	0.75	0.61
item 5	0.72	0.70	0.77	0.74	1	0.83	0.70
item 6	0.75	0.75	0.83	0.78	0.85	1	0.70
item 7	0.66	0.65	0.69	0.67	0.71	0.78	1

study 2	**item 1**	**item 2**	**item 3**	**item 4**	**item 5**	**item 6**	**item 7**
item 1	1	0.77	0.78	0.75	0.75	0.74	0.64
item 2	0.76	1	0.82	0.76	0.75	0.75	0.63
item 3	0.80	0.81	1	0.79	0.79	0.80	0.65
item 4	0.78	0.75	0.80	1	0.78	0.79	0.62
item 5	0.78	0.76	0.79	0.77	1	0.83	0.68
item 6	0.75	0.77	0.81	0.78	0.82	1	0.71
item 7	0.67	0.71	0.66	0.67	0.68	0.72	1

Table 3.3: Correlations among actors across all items, for perspectives for Others (above the diagonal) and for Yourself (below the diagonal).

study 1	**Nature**	**God**	**Institutions**	**Other People**	**Yourself**	**Chance**
Nature	1	0.36	0.23	0.24	0.20	0.29
God	0.20	1	0.01	-0.09	-0.00	-0.12
Institutions	0.27	0.02	1	0.65	0.26	0.16
Other People	0.14	-0.16	0.66	1	0.29	0.31
Yourself	0.09	-0.06	0.14	0.20	1	0.15
Chance	0.36	-0.17	0.23	0.30	0.12	1
study 2	**Nature**	**God**	**Institutions**	**Other People**	**Yourself**	**Chance**
Nature	1	0.34	0.31	0.29	0.19	0.33
God	0.26	1	0.09	0.02	-0.05	-0.10
Institutions	0.29	0.07	1	0.71	0.35	0.22
Other People	0.34	0.03	0.73	1	0.38	0.25
Yourself	0.18	-0.15	0.25	0.23	1	0.25
Chance	0.38	-0.13	0.20	0.22	0.25	1

In both studies, the actor correlations are fairly low except between Human Institutions and Other People.

3.3.2 Preliminary analysis of t-tests to compare means

To compare the means for the two studies, we do a two-sample t-test for each item and actor and perspective separately. We assume that the variances are equal in the two samples (but this is questionable). Hence, we obtain 84 t-values in total as in Table 3.4. There are interesting patterns of large t-values (both positive and negative), which are quite similar for the two perspectives. Several methods have been proposed to control the family wise error rate in case of multiple comparisons. The simple and conservative Bonferroni correction implies replacing α by $\alpha/84$ in our case. This results in critical values ± 3.43 and yields 1 significant t-value for the Others perspective, and also 1 significant t-value for the Yourself perspective. The Bonferroni-Holm procedure (less conservative) orders the t-values and starts with the largest in magnitude and $\alpha/84$, where 84 is decreased by one for each null hypothesis that is rejected. This procedure also results in 1 significant t-value for both perspectives. We conclude that overall the mean differences between the two studies are rather small with only 2 out of 84 t-values being significant.

Table 3.4: T-values of two-sample t-tests per item and actor and perspective (study 1 minus study 2), to compare means for the two studies. Values in bold are significant at 5% level (two-sided, critical values are ±1.96).

for Others	**item 1**	**item 2**	**item 3**	**item 4**	**item 5**	**item 6**	**item 7**
Nature	0.83	0.77	0.70	1.30	0.75	0.33	0.27
God	**-2.96**	**-3.11**	**-3.61**	**-2.41**	**-3.08**	**-2.90**	**-2.86**
Institutions	0.59	0.35	0.86	0.10	-0.08	0.12	-0.69
Other People	1.10	0.38	1.75	1.85	0.86	1.45	-0.54
Yourself	**2.04**	1.73	**2.40**	**2.53**	1.77	1.61	1.36
Chance	0.79	-0.44	0.17	1.14	-0.23	0.20	0.49
for Yourself	**item 1**	**item 2**	**item 3**	**item 4**	**item 5**	**item 6**	**item 7**
Nature	1.26	1.39	1.30	0.61	1.74	2.34	1.99
God	**-2.45**	**-2.09**	**-2.34**	**-2.14**	**-2.16**	**-2.20**	-1.78
Institutions	1.72	**2.37**	1.78	1.80	1.06	1.92	**2.32**
Other People	**2.96**	**3.61**	**2.28**	**3.28**	1.51	1.92	**2.14**
Yourself	0.43	0.39	1.68	1.94	**2.30**	**2.35**	**3.17**
Chance	0.23	-0.39	0.97	-0.22	-0.01	0.28	0.85

To compare means for the two perspectives, we also do a paired sample t-test for each item and actor separately for the two studies. Hence, we obtain 42 t-values for each study. In Table 3.5, there are interesting patterns of large t-values (both positive and negative), which are quite different for the two studies. The Bonferroni correction implies replacing α by $\alpha/42$ in our case. This results in critical values ± 3.24 and yields 3 significant t-values for study 1, and also 3 significant t-values for study 2. The Bonferroni-Holm procedure also results in 3 significant t-values for both studies. We conclude that overall the mean differences between the two perspectives are rather small with only 3 out of 42 t-values being significant in both studies.

Table 3.5: T-values of paired sample t-tests per item and actor (for Others minus for Yourself), to compare means for the two perspectives. Values in bold are significant at 5% level (two-sided, critical values are ±1.96).

study 1	**item 1**	**item 2**	**item 3**	**item 4**	**item 5**	**item 6**	**item 7**
Nature	0.48	0.28	0.19	1.39	0.39	-1.44	**-2.91**
God	-0.07	0.19	-0.47	0.21	-0.44	-0.59	-1.11
Institutions	-0.85	-0.69	0.14	0.77	1.21	-0.41	**-5.19**
Other People	-1.43	**-2.93**	0.26	0.00	**2.53**	0.17	**-5.46**
Yourself	1.69	1.07	0.14	1.38	**-2.09**	**-2.08**	**-3.27**
Chance	1.19	1.16	**2.71**	**2.71**	1.21	0.53	**-1.97**
study 2	**item 1**	**item 2**	**item 3**	**item 4**	**item 5**	**item 6**	**item 7**
Nature	1.60	1.90	1.64	0.42	**2.69**	**3.04**	-0.43
God	0.85	**2.78**	**2.69**	0.66	**2.00**	1.10	1.54
Institutions	0.83	**2.74**	**1.98**	**4.26**	**3.89**	**2.87**	-1.22
Other People	1.37	1.81	1.73	**2.87**	**4.76**	1.23	**-2.36**
Yourself	-0.18	-0.85	-1.10	0.88	-1.80	-1.29	-1.29
Chance	0.47	1.32	0.89	0.67	1.92	0.72	**-2.03**

In each study, the order of the questions is varied, with order= 1 for first Others and then Yourself, order= 2 for first Yourself and then Others. The numbers of subjects (without missing data) with the same order of questions are as follows

study 1:	61 with order= 1,	48 with order= 2
study 2:	118 with order= 1,	118 with order= 2

We do a two-sample t-test for each item and actor and perspective and study (order= 1 minus order= 2) to compare means for the orders of questions for every study.

For study 1, the t-values of perspective "for Others" are significant at the 5% level (two-sided, critical values are ± 1.96) on items 3 and 5 of actor Institutions (2.26 and 1.97), and on item 1 of actor Other People (-2.12). For perspective "for Yourself", the t-values are significant at the 5% level (two-sided, critical values are ± 1.96) on items 1 and 2 of actor Nature (-2.14 and -2.33).

For study 2, the t-values for perspective "for Others" are significant at the 5% level (two-sided, critical values are ± 1.96) on item 1 of actor God (-2.37), on items 4 and 6 of actor Chance (2.32 and 2.10). The t-values for perspective "for Yourself" are significant at the 5% level (two-sided, critical values are ± 1.96) on item 2 of actor Yourself (2.42), and on item 7 of actor Chance (2.74).

The Bonferroni correction (critical values ± 3.43) and the Bonferroni-Holmes procedure both yield no significant t-values out of 84. We conclude that overall the mean differences between the two orders are not significant for both studies 1 and 2.

3.4 PCA for study 1 and 2

For each study and each perspective separately, we do a PCA with Varimax rotation (resulting in (hopefully) interpretable orthogonal components) on the

correlation matrix with 42 items (7 items for each of the 6 actors). When $\mathbf{X}_{kl}$ denotes the matrix of scores of size subjects $\times$ items for actor k and perspective l, this implies that a PCA is done on unfolded data $[\mathbf{X}_{1l} \ldots \mathbf{X}_{6l}]$ of size $N \times 42$. We use 5 components in each PCA. Adding a sixth component yields either a non-interpretable component or a component with small loadings (around 0.4).

The explained variances of the PCAs with 5 components are good (70%-75%). After rotation, the five components have a very clear interpretation: each component is interpreted as one actor, except for Institutions and Other People. These are found combined in one factor. This reflects the high correlations between these actors and the low correlations between the other actors. The loadings of some components are less pronounced for item 7, which correlates the least with other items. The eigenvalues of the sixth components are around 1-1.5, while the eigenvalues of the fifth components are around 2.1-2.5. The eigenvalues larger than 1 criterion indicates that we should include 5 (or 6) components for each PCA. The PCA results featuring 5 actors are conform the common factor analysis results obtained in Stroebe et al. (2014). The 5 extracted actors are the same as in the new BJW scale proposed by Stroebe et al. (2014).

Since the component structure is almost the same in study 1 and study 2, we combine both studies in one dataset and do a PCA with Varimax rotation for each perspective separately. The results are presented in Tables 3.6, 3.8, and **??**. We use Varimax rotation with Kaiser normalization. These solutions are analogous to the results of the two studies separately. Therefore, the conclusions are also the same as for the PCAs of the two studies separately.

Table 3.6: Explained variance in PCA on the unfolded data per perspective, for both studies together

for Others					
Component	Extraction Sums of Squared Loading		Rotation Sums of Squared Loading		Interpretation
	Eigenvalue	% of Variance	Eigenvalue	% of Variance	
1	12.7	30.3	8.3	19.9	Inst.-Other People
2	7.7	18.2	6.5	15.5	God
3	4.6	10.9	5.4	12.8	Nature
4	3.4	8.0	5.3	12.6	Yourself
5	2.2	5.1	4.9	11.7	Chance
Total	30.6	72.5	30.4	72.5	
for Yourself					
Component	Extraction Sums of Squared Loading		Rotation Sums of Squared Loading		Interpretation
	Eigenvalue	% of Variance	Eigenvalue	% of Variance	
1	12.3	29.4	8.9	21.1	Inst-Other People
2	7.5	17.9	6.5	15.5	God
3	5.2	12.3	5.5	13.2	Nature
4	3.8	9.1	5.2	12.5	Yourself
5	2.4	5.8	5.1	12.0	Chance
Total	31.2	74.5	31.2	74.3	

Table 3.7: Rotated PCA Component matrix for Others

actor	item	Component				
		1	2	3	4	5
Nature	item 1	0.22	0.16	**0.80**	0.10	0.16
	item 2	0.15	0.23	**0.82**	0.10	0.12
	item 3	0.15	0.21	**0.84**	0.09	0.19
	item 4	0.18	0.18	**0.83**	0.07	0.23
	item 5	0.14	0.25	**0.81**	0.09	0.17
	item 6	0.18	0.21	**0.85**	0.12	0.18
	item 7	0.23	0.12	**0.72**	0.14	0.06
God	item 1	0.00	**0.93**	0.18	-0.01	-0.07
	item 2	0.00	**0.94**	0.17	-0.05	-0.08
	item 3	0.01	**0.94**	0.18	-0.04	-0.07
	item 4	-0.01	**0.92**	0.19	-0.03	-0.07
	item 5	-0.03	**0.94**	0.16	-0.05	-0.09
	item 6	0.01	**0.94**	0.18	-0.00	-0.07
	item 7	-0.01	**0.90**	0.15	-0.02	-0.12
Institutions	item 1	**0.73**	0.01	0.11	0.12	0.16
	item 2	**0.77**	0.05	0.03	0.09	0.07
	item 3	**0.76**	0.07	0.02	0.13	0.05
	item 4	**0.72**	0.00	0.05	0.08	0.14
	item 5	**0.77**	0.11	0.15	0.17	0.04
	item 6	**0.80**	0.06	0.16	0.18	-0.00
	item 7	**0.66**	0.16	0.16	0.27	-0.05
Other People	item 1	**0.70**	-0.05	0.16	0.14	0.20
	item 2	**0.74**	-0.10	0.08	0.04	0.17
	item 3	**0.76**	-0.07	0.11	0.16	0.07
	item 4	**0.76**	-0.11	0.07	0.11	0.17
	item 5	**0.73**	-0.05	0.22	0.20	0.10
	item 6	**0.74**	-0.09	0.22	0.23	0.04
	item 7	**0.66**	0.04	0.16	0.29	-0.02
Yourself	item 1	0.18	0.01	0.09	**0.79**	0.07
	item 2	0.17	-0.10	0.09	**0.79**	0.13
	item 3	0.22	-0.01	0.09	**0.84**	0.08
	item 4	0.22	-0.06	0.06	**0.84**	0.12
	item 5	0.24	-0.04	0.06	**0.85**	0.04
	item 6	0.28	-0.03	0.09	**0.85**	0.04
	item 7	0.24	0.02	0.18	**0.79**	0.07
Chance	item 1	0.09	-0.11	0.11	0.07	**0.82**
	item 2	0.09	-0.11	0.12	0.06	**0.85**
	item 3	0.08	-0.11	0.12	0.09	**0.88**
	item 4	0.15	-0.07	0.11	-0.00	**0.83**
	item 5	0.12	-0.10	0.18	0.06	**0.88**
	item 6	0.16	-0.00	0.24	0.18	**0.80**
	item 7	0.27	-0.02	0.32	0.23	**0.42**

Table 3.8: Rotated PCA Component matrix for Yourself

actor	item	Component				
		1	2	3	4	5
Nature	item 1	0.17	0.13	**0.81**	0.08	0.21
	item 2	0.17	0.19	**0.87**	0.12	0.13
	item 3	0.10	0.12	**0.88**	0.07	0.19
	item 4	0.13	0.13	**0.81**	0.06	0.26
	item 5	0.14	0.15	**0.86**	0.07	0.20
	item 6	0.17	0.14	**0.88**	0.08	0.14
	item 7	0.31	0.05	**0.72**	0.07	0.05
God	item 1	-0.02	**0.94**	0.14	-0.07	-0.06
	item 2	0.00	**0.94**	0.12	-0.06	-0.10
	item 3	-0.02	**0.94**	0.14	-0.09	-0.08
	item 4	-0.04	**0.94**	0.12	-0.09	-0.06
	item 5	-0.03	**0.95**	0.11	-0.07	-0.10
	item 6	0.00	**0.95**	0.13	-0.07	-0.10
	item 7	0.05	**0.92**	0.09	-0.06	-0.12
Institutions	item 1	**0.75**	0.03	0.05	0.08	0.12
	item 2	**0.81**	0.06	0.04	0.10	0.00
	item 3	**0.79**	0.02	0.03	0.13	0.14
	item 4	**0.73**	-0.00	0.10	0.06	0.16
	item 5	**0.76**	0.12	0.20	0.12	0.03
	item 6	**0.83**	0.06	0.08	0.15	0.06
	item 7	**0.80**	0.04	0.07	0.16	-0.06
Other People	item 1	**0.72**	-0.04	0.13	0.05	0.09
	item 2	**0.75**	-0.05	0.09	0.03	-0.02
	item 3	**0.76**	-0.10	0.15	0.07	0.15
	item 4	**0.75**	-0.08	0.09	0.08	0.19
	item 5	**0.73**	0.00	0.26	0.11	0.11
	item 6	**0.81**	-0.08	0.12	0.20	0.09
	item 7	**0.76**	-0.03	0.08	0.20	-0.04
Yourself	item 1	0.08	-0.00	-0.02	**0.85**	0.08
	item 2	0.10	-0.02	0.05	**0.77**	0.13
	item 3	0.18	-0.08	0.13	**0.83**	0.03
	item 4	0.19	-0.08	0.12	**0.86**	0.12
	item 5	0.19	-0.18	0.02	**0.85**	0.13
	item 6	0.21	-0.08	0.12	**0.86**	0.01
	item 7	0.18	-0.07	0.14	**0.80**	0.10
Chance	item 1	0.13	-0.11	0.14	0.09	**0.84**
	item 2	0.00	-0.07	0.09	0.08	**0.87**
	item 3	0.05	-0.08	0.19	0.05	**0.88**
	item 4	0.14	-0.12	0.10	0.09	**0.84**
	item 5	0.10	-0.14	0.19	0.07	**0.89**
	item 6	0.17	-0.04	0.29	0.15	**0.76**
	item 7	0.30	-0.09	0.32	0.20	**0.50**

3.5 Tucker3 for studies 1 and 2 together

As an alternative to PCA perspective as above, we could also fit the three-way CP model to the $345 \times 7 \times 6$ array for each perspective. However, CP (per perspective) does not yield one actor per component as in PCA with Varimax rotation. In this section we fit the Tucker3 model with Joint Orthomax rotation of Kiers (1998a) for each perspective and obtain one actor per component. To fit the Tucker3 models, we use an alternating least squares algorithm (Kroonenberg & De Leeuw, 1980) programmed in Matlab by Henk A.L. Kiers (available at http://www.gmw.rug.nl/∼kiers/). We use 10 random starts and convergence criterion $1e-9$.

Let $\underline{\mathbf{X}}_1$ and $\underline{\mathbf{X}}_2$ be two $345 \times 7 \times 6$ arrays containing the data of perspective "for Others" and "for Yourself", respectively. $\underline{\mathbf{X}}_1$ and $\underline{\mathbf{X}}_2$ are centered across subjects (i.e. the columns of $\underline{\mathbf{X}}_1$ and $\underline{\mathbf{X}}_2$ are centered). We fit the Tucker3 model for both $\underline{\mathbf{X}}_1$ and $\underline{\mathbf{X}}_2$ with R subject components, P item components, and 5 actor components ($R, P \in \{2, \ldots, 5\}$). We choose 5 actor components because we hope to find one actor per component as in PCA with Varimax rotation. Table 3.9 shows the fit percentages for each case of $(R, P, 5)$.

Table 3.9: Tucker3 fit percentages for each case of $(R, P, 5)$

	for Others					for Yourself				
R╲ P	1	2	3	4	5	1	2	3	4	5
2	53.99	54.20	54.26	54.30	54.33	52.51	52.61	52.71	52.75	52.77
3	63.39	63.86	63.94	63.99	64.03	62.75	63.16	63.28	63.37	63.41
4	69.55	70.06	70.15	70.21	70.25	70.55	71.00	71.13	71.22	71.28
5	75.02	75.65	75.77	75.83	75.87	76.40	76.99	77.14	77.27	77.36

Since fit percentages are not much different for various values of P but much increasing when R increases, we choose $R = 5$ and $P = 1$ for both perspectives. In the rotation method of Kiers (1998a), the rotation criterion consists of one term for each mode and one term for the core array. Each term expresses simple structure in a mode or the core array, and has a separate weight. We are only interested in obtaining simple structure in the actor mode and the core array, which we give equal weight. The weights for the person and item modes are set to zero. The rotated Tucker3 solution with $R = 5$, $P = 1$ for others is as follows

$$\mathbf{B} = \begin{pmatrix} 0.94 \\ 1.00 \\ 1.01 \\ 0.99 \\ 1.06 \\ 1.04 \\ 0.96 \end{pmatrix} \begin{matrix} \text{Item 1} \\ \text{Item 2} \\ \text{Item 3} \\ \text{Item 4} \\ \text{Item 5} \\ \text{Item 6} \\ \text{Item 7} \end{matrix}, \quad \mathbf{C} = \begin{pmatrix} -0.02 & 0.46 & 0.30 & 0.01 & \mathbf{2.39} \\ 0.05 & \mathbf{2.40} & -0.10 & 0.11 & -0.44 \\ -0.27 & 0.13 & \mathbf{1.75} & -0.21 & -0.31 \\ -0.21 & -0.05 & \mathbf{1.63} & -0.11 & -0.13 \\ \mathbf{2.42} & -0.03 & 0.35 & -0.01 & -0.02 \\ -0.02 & -0.10 & 0.23 & \mathbf{2.44} & -0.02 \end{pmatrix} \begin{matrix} \text{Nature} \\ \text{God} \\ \text{Inst.} \\ \text{Other} \\ \text{Yourself} \\ \text{Chance} \end{matrix} \tag{3.1}$$

The frontal slices of the core array are shown in Table 3.10 with the percentage of explained variance per core entry parenthesized. The latter are possible to compute due to the orthogonality of the components in each of the three modes. The interpretation of (3.1) is as follows. Since each $\mathbf{G}_q$ has only one value that is not close to zero (see Table 3.10), the Tucker3 model with $(5, 1, 5)$ components

$$\sum_{r=1}^{5} \sum_{q=1}^{5} g_{r1q} (\mathbf{a}_r \circ \mathbf{B} \circ \mathbf{c}_q)$$

has only 5 large terms, each term corresponding to one actor. Term q is corresponding to column q of $\mathbf{C}$. Hence, component 1 is interpreted as Yourself,

component 2 as God, component 3 as Institutions and Other People, component 4 as Chance, and component 5 as Nature.

Table 3.10: The frontal slices of the core "for Others" array with the percentage of explained variance per core entry parenthesized

	$\mathbf{G}_1$ (Yourself)	$\mathbf{G}_2$ (God)	$\mathbf{G}_3$ (Inst.-Other)	$\mathbf{G}_4$ (Chance)	$\mathbf{G}_5$ (Nature)
Comp 1	-0.02 (0.00)	**1.64 (28.39)**	0.04 (0.02)	-0.05 (0.03)	0.21 (0.44)
Comp 2	0.05 (0.02)	-0.01 (0.00)	0.11 (0.12)	**0.94 (9.27)**	0.20 (0.41)
Comp 3	0.06 (0.04)	0.04 (0.02)	0.12 (0.15)	0.18 (0.34)	**0.91 (8.68)**
Comp 4	**0.86 (7.85)**	0.00 (0.00)	0.11 (0.13)	0.04 (0.02)	0.05 (0.03)
Comp 5	0.32 (1.09)	0.02 (0.01)	**1.26 (16.54)**	0.23 (0.57)	0.29 (0.89)

The rotated Tucker3 solution with $R = 5$, $P = 1$ for Yourself is as follows

$$\mathbf{B} = \begin{pmatrix} 0.98 \\ 0.99 \\ 1.00 \\ 1.00 \\ 1.05 \\ 1.02 \\ 0.95 \end{pmatrix} \begin{matrix} \text{Item 1} \\ \text{Item 2} \\ \text{Item 3} \\ \text{Item 4} \\ \text{Item 5} \\ \text{Item 6} \\ \text{Item 7} \end{matrix}, \quad \mathbf{C} = \begin{pmatrix} 0.20 & -0.02 & \mathbf{2.43} & -0.04 & 0.23 \\ 0.04 & 0.16 & -0.23 & 0.18 & \mathbf{2.42} \\ \mathbf{1.76} & -0.14 & -0.23 & -0.12 & 0.05 \\ \mathbf{1.67} & -0.16 & -0.05 & -0.12 & -0.10 \\ 0.21 & \mathbf{2.44} & 0.02 & -0.01 & -0.16 \\ 0.17 & -0.01 & 0.05 & \mathbf{2.44} & -0.18 \end{pmatrix} \begin{matrix} \text{Nature} \\ \text{God} \\ \text{Inst.} \\ \text{Other} \\ \text{Yourself} \\ \text{Chance} \end{matrix} \tag{3.2}$$

The frontal slices of the core array are shown in Table 3.11 with the percentage of explained variance per core entry parenthesized. The interpretation of (3.2) is as follows. In Table 3.11, each $\mathbf{G}_q$ has only one value that is not close to zero. This implies we obtain one actor per component. Component 1 is interpreted as Institutions and Other People, component 2 as Yourself, component 3 as Nature, component 4 as Chance, and component 5 as God. Compare to

(3.1), this solution is very similar except for the order of components.

Table 3.11: The frontal slices of the core array "for Yourself" with the percentage of explained variance per core entry parenthesized

	$\mathbf{G}_1$ (Inst.-Other)	$\mathbf{G}_2$ (Yourself)	$\mathbf{G}_3$ (Nature)	$\mathbf{G}_4$ (Chance)	$\mathbf{G}_5$ (God)
Comp 1	0.11 (0.12)	**0.97 (9.18)**	0.07 (0.05)	0.06 (0.04)	-0.02 (0.01)
Comp 2	-0.02 (0.00)	-0.09 (0.08)	0.13 (0.18)	-0.10 (0.10)	**1.66 (26.98)**
Comp 3	0.09 (0.08)	0.07 (0.04)	0.20 (0.39)	**0.97 (9.28)**	-0.02 (0.01)
Comp 4	0.16 (0.25)	0.09 (0.09)	**1.06 (10.98)**	0.25 (0.63)	0.04 (0.02)
Comp 5	**1.29 (16.28)**	0.24 (0.54)	0.27 (0.74)	0.19 (0.35)	-0.01 (0.00)

Since the item weights in $\mathbf{B}$ are nearly identical, the Tucker3 solutions (3.1) and (3.2) are very similar to the PCA solutions in section 3.4.

3.6 Tucker4 for both perspectives together

Since the Tucker3 solutions for the two perspectives are very similar, in this section we try Tucker4 for the 4-way BJW dataset of size $345 \times 7 \times 6 \times 2$. The Tucker4 model is defined analogous to the Tucker3 model in (2.9). Let $\underline{\mathbf{X}}$ be a 4-way array of size $I \times J \times K \times L$. The Tucker4 model is given by (e.g., Kroonenberg, 2008)

$$\underline{\mathbf{X}} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} \sum_{s=1}^{S} g_{pqrs} (\mathbf{a}_p \circ \mathbf{b}_q \circ \mathbf{c}_r \circ \mathbf{d}_s) + \underline{\mathbf{E}}, \tag{3.3}$$

where "$\circ$" denotes outer product of vectors, $\mathbf{a}_p$, $\mathbf{b}_q$, $\mathbf{c}_r$, $\mathbf{d}_s$ are columns of component matrices $\mathbf{A}$ ($I \times P$), $\mathbf{B}$ ($J \times Q$), $\mathbf{C}$ ($K \times R$), and $\mathbf{D}$ ($L \times S$), respectively, g_{pqrs} is an element of the $P \times Q \times R \times S$ core array $\underline{\mathbf{G}}$, and $\mathbf{E}$ is the array of residuals. To fit the Tucker4 model, we extended the alternating least squares

algorithm programmed in Matlab by Henk A.L. Kiers to four modes. We use 10 random starts and convergence criterion $1e-9$.

3.6.1 The Tucker4 solution

Table 3.12 shows the fit percentages for each case of $(R, P, 5, Q)$, where $R = 2, \ldots, 4$, $P = 1, 2$, and $Q = 1, 2$:

Table 3.12: Tucker4 fit percentages for each case of $(R, P, 5, Q)$

$Q = 1$			$Q = 2$		
R\ P	1	2	R\ P	1	2
2	49.18	49.32	2	49.25	49.41
3	57.89	58.22	3	57.98	58.35
4	63.81	64.18	4	64.06	64.47
5	68.87	69.35	5	69.14	69.68

The fit percentages are not much different between $Q = 1$ and $Q = 2$, and between $P = 1$ and $P = 2$. However, they are much increasing when R increases. Therefore, we choose $Q = P = 1$, and $R = 5$. We extended the Tucker3 rotation method of Kiers (1998a) to the Tucker4 model; see section 3.6.2. We are only interested in obtaining simple structure in the actor mode and the core array. The weights for the person, item, and perspective modes are set to zero. The rotated Tucker4 solution with 5 subject components, 1 item component, 5 actor components, and 1 perspective component is:

$$\mathbf{B} = \begin{pmatrix} 0.96 \\ 0.98 \\ 1.00 \\ 1.00 \\ 1.06 \\ 1.04 \\ 0.96 \end{pmatrix} \begin{matrix} \text{Item 1} \\ \text{Item 2} \\ \text{Item 3} \\ \text{Item 4} \\ \text{Item 5} \\ \text{Item 6} \\ \text{Item 7} \end{matrix}, \quad \mathbf{C} = \begin{pmatrix} 0.16 & \mathbf{2.42} & -0.04 & -0.02 & 0.33 \\ 0.04 & -0.32 & 0.17 & 0.12 & \mathbf{2.42} \\ \mathbf{1.76} & -0.19 & -0.14 & -0.13 & 0.05 \\ \mathbf{1.67} & -0.03 & -0.07 & -0.19 & -0.11 \\ 0.23 & 0.03 & -0.01 & \mathbf{2.44} & -0.12 \\ 0.14 & 0.05 & \mathbf{2.44} & -0.01 & -0.16 \end{pmatrix} \begin{matrix} \text{Nature} \\ \text{God} \\ \text{Inst.} \\ \text{Other} \\ \text{Yourself} \\ \text{Chance} \end{matrix},$$

$$\mathbf{D} = \begin{pmatrix} 0.97 \\ 1.03 \end{pmatrix} \begin{matrix} \text{Others} \\ \text{Yourself} \end{matrix}. \tag{3.4}$$

The frontal slices of the core array are shown in Table 3.13 with the percentage of explained variance per core entry parenthesized.

Table 3.13: The frontal slices of the core array with the percentage of explained variance per core entry parenthesized

	$\mathbf{G}_1$ (Inst.-Other)	$\mathbf{G}_2$ (Nature)	$\mathbf{G}_3$ (Chance)	$\mathbf{G}_4$ (Yourself)	$\mathbf{G}_5$ (God)
Comp 1	0.17 (0.30)	**0.98 (9.76)**	0.25 (0.61)	0.10 (0.11)	0.04 (0.02)
Comp 2	-0.02 (0.00)	0.16 (0.25)	-0.08 (0.06)	-0.06 (0.03)	**1.60 (25.74)**
Comp 3	0.10 (0.10)	0.19 (0.37)	**0.90 (8.26)**	0.07 (0.05)	-0.02 (0.00)
Comp 4	**1.15 (13.41)**	0.27 (0.72)	0.19 (0.36)	0.28 (0.79)	-0.01 (0.00)
Comp 5	0.13 (0.16)	0.08 (0.06)	0.06 (0.04)	**0.87 (7.67)**	-0.01 (0.00)

The interpretation of (3.4) is as follows. Since each $\mathbf{G}_q$ has only one value that is not close to zero (see Table 3.13), the Tucker4 model with $(5, 1, 5, 1)$ components

$$\sum_{r=1}^{5} \sum_{q=1}^{5} g_{r1q1} (\mathbf{a}_r \circ \mathbf{B} \circ \mathbf{c}_q \circ \mathbf{D})$$

has only 5 large terms, term q corresponding to column q of matrix $\mathbf{C}$ in (3.4). This implies we obtain one actor per component. Component 1 represent a combination of Institutions and Other People. Component 2 is interpreted as Nature, component 3 as Chance, component 4 as Yourself, and component 5 as God.

Since weights in $\mathbf{B}$ are nearly identical and those in $\mathbf{D}$ too, the Tucker4 solution is very similar to the PCA solution in section 3.4. Compared to the solutions in section 3.5, this solution is very similar except for the order of component.

3.6.2 The Tucker4 rotation

The rotation we used here is an extension of Kiers (1998a). First, we consider the Tucker4 model as defined by (3.3). We can also write the Tucker4 model by using Kronecker products. Matricizing $\underline{\mathbf{X}}$, $\underline{\mathbf{G}}$, $\underline{\mathbf{E}}$ as $\mathbf{X}_a = [\mathbf{X}_1|\cdots|\mathbf{X}_L]$, $\mathbf{G}_a = [\mathbf{G}_1|\cdots|\mathbf{G}_S]$, and $\mathbf{E}_a = [\mathbf{E}_1|\cdots|\mathbf{E}_L]$, where $\mathbf{X}_l$ $(I \times JK)$ is the unfolding of three-way array $\underline{\mathbf{X}}(:,:,:,l)$ $(I \times J \times K)$ into its frontal slices next to each other, $\mathbf{E}_l$ $(I \times JK)$ is the unfolding of three-way array $\underline{\mathbf{E}}(:,:,:,l)$ $(I \times J \times K)$ into its frontal slices next to each other, and $\mathbf{G}_s$ $(P \times QR)$ is the unfolding of three-way array $\underline{\mathbf{G}}(:,:,:,s)$ $(P \times Q \times R)$ into its frontal slices next to each other, we have

$$\mathbf{X}_a = \mathbf{A}\mathbf{G}_a\left(\mathbf{D}^T \otimes (\mathbf{C}^T \otimes \mathbf{B}^T)\right) + \mathbf{E}_a. \tag{3.5}$$

Analogous to Tucker3, a Tucker4 solution is not unique. Postmultiplying $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ by non-singular matrices $\mathbf{T}$, $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$ can always be compensated for by applying the inverses of these matrices to the core array $\underline{\mathbf{G}}$. That is $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{T}^{-1}$, $\tilde{\mathbf{B}} = \mathbf{B}\mathbf{U}^{-1}$, $\tilde{\mathbf{C}} = \mathbf{C}\mathbf{V}^{-1}$, $\tilde{\mathbf{D}} = \mathbf{D}\mathbf{W}^{-1}$, and $\underline{\tilde{\mathbf{G}}}$ with elements

$$\tilde{g}_{ijkl} = \sum_{p=1}^{P}\sum_{q=1}^{Q}\sum_{r=1}^{R}\sum_{s=1}^{S} t_{ip}u_{jq}v_{kr}w_{ls}g_{pqrs}, \tag{3.6}$$

$i = 1 \ldots P$, $j = 1 \ldots Q$, $k = 1 \ldots R$, $l = 1 \ldots S$, give the same model estimates as $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, and $\underline{\mathbf{G}}$ do. In other words, $\tilde{\mathbf{A}} = \mathbf{A}\mathbf{T}^{-1}$, $\tilde{\mathbf{B}} = \mathbf{B}\mathbf{U}^{-1}$, $\tilde{\mathbf{C}} = \mathbf{C}\mathbf{V}^{-1}$, $\tilde{\mathbf{D}} = \mathbf{D}\mathbf{W}^{-1}$, and

$$\tilde{\mathbf{G}}_a = \mathbf{T}\mathbf{G}_a \left(\mathbf{W}^T \otimes (\mathbf{V}^T \otimes \mathbf{U}^T)\right) \tag{3.7}$$

give the same model estimates as $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, and $\mathbf{G}_a$ do.

Therefore, to get a simple structure in the core, we can extend the Joint Orthomax rotation of Kiers (1998a) for Tucker3 to Tucker4. This method is based on choosing different sets of weights for component matrices to obtain different tradeoffs between simplicity of the core and simplicity of the component matrices. Next, we explain how the Joint Orthomax criterion can be optimized over $\mathbf{T}$, $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$.

Let $\mathbf{\Lambda}$ be a $M \times N$ matrix. The orthomax criterion (Crawford & Ferguson, 1970; Jennrich, 1970), which has varimax and quartimax as special cases, is given by $M^{-1} f_{OR}$, where f_{OR} is defined as

$$f_{OR}(\mathbf{\Lambda}, \gamma) = \sum_{n=1}^{N} \left[\sum_{m=1}^{M} \lambda_{mn}^4 - \frac{\gamma}{M} \left[\sum_{m=1}^{M} \lambda_{mn}^2 \right]^2 \right], \tag{3.8}$$

where λ_{mn} stands for the element $(m, n)$of $\mathbf{\Lambda}$, and γ is the parameter monitoring the choice of orthomax criterion (e.g., $\gamma = 0$ yields the quartimax criterion, and $\gamma = 1$ yields the varimax criterion). Then the four-mode orthomax criterion is given by

$$f(\mathbf{T}, \mathbf{U}, \mathbf{V}, \mathbf{W}) = \sum_{n=1}^{4} w_n f_{OR}(\tilde{\mathbf{G}}^n, \gamma_n), \tag{3.9}$$

where $\tilde{\mathbf{G}}^n$ denotes the matrix that is obtained by matricizing the three-way arrays with the n-th index of $\underline{\tilde{\mathbf{G}}}$ being fixed; $\underline{\tilde{\mathbf{G}}}$ is the original core array rotated by the orthonormal matrices $\mathbf{T}$, $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$, according to (3.6); w_1, w_2, w_3, and w_4 are fixed prespecified weights; and γ_1, γ_2, γ_3, and γ_4 are prespecified values of the γ parameter in (3.8).

The joint orthomax criterion for the core and the component matrices is given by

$$f_{cc}(\mathbf{T},\mathbf{U},\mathbf{V},\mathbf{W}) = \sum_{n=1}^{4} w_n f_{OR}(\tilde{\mathbf{G}}^n,\gamma_n) + w_A f_{OR}(\tilde{\mathbf{A}},\gamma_A) + w_B f_{OR}(\tilde{\mathbf{B}},\gamma_B) + w_C f_{OR}(\tilde{\mathbf{C}},\gamma_C) + w_D f_{OR}(\tilde{\mathbf{D}},\gamma_D), \tag{3.10}$$

where w_A, w_B, w_C, and w_D denote the weights attached to the orthomax criteria applied respectively to $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, $\tilde{\mathbf{C}}$, and $\tilde{\mathbf{D}}$, with respective γ-parameters γ_A, γ_B, γ_C, and γ_D. Joint orthomax rotation of the core and the component matrices consists of maximizing f_{cc} for a priori chosen weights w_1, w_2, w_3, w_4, w_A, w_B, w_C, and w_D, and a priori chosen orthomax parameters γ_1, γ_2, γ_3, γ_4, γ_A, γ_B, γ_C, and γ_D, with $\tilde{\underline{\mathbf{G}}}$ as in (3.6), and $\tilde{\mathbf{A}}$, $\tilde{\mathbf{B}}$, $\tilde{\mathbf{C}}$, and $\tilde{\mathbf{D}}$ as defined above, over orthonormal matrices $\mathbf{T}$, $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{W}$.

The weights and the orthomax parameters in the joint orthomax criterion can be specified at will. However, in practice, some guidelines for these choices may be helpful. In sections 3.5 and 3.6.1 we use the varimax rotation criterion only. Hence,

$$\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_{\mathbf{A}} = \gamma_{\mathbf{B}} = \gamma_{\mathbf{C}} = \gamma_{\mathbf{D}} = 1.$$

For the weights, we use the following "standard" weights, which are defined analogous to Kiers (1998a):

$$w_1 = \nu_1(\nu_1+\nu_2+\nu_3+\nu_4)^{-1}(PQRS)||\underline{\mathbf{G}}||^4, \text{ where } \nu_1 = (QRS)^{-1} \text{ or } \nu_1 = 0$$

$$w_2 = \nu_2(\nu_1+\nu_2+\nu_3+\nu_4)^{-1}(PQRS)||\underline{\mathbf{G}}||^4, \text{ where } \nu_2 = (PRS)^{-1} \text{ or } \nu_2 = 0$$

$$w_3 = \nu_3(\nu_1+\nu_2+\nu_3+\nu_4)^{-1}(PQRS)||\underline{\mathbf{G}}||^4, \text{ where } \nu_3 = (PQS)^{-1} \text{ or } \nu_3 = 0$$

$$w_4 = \nu_4(\nu_1+\nu_2+\nu_3+\nu_4)^{-1}(PQRS)||\underline{\mathbf{G}}||^4, \text{ where } \nu_4 = (PQR)^{-1} \text{ or } \nu_4 = 0$$

$$w_A = IP^{-1}; \quad w_B = JQ^{-1}; \quad w_C = KR^{-1}; \quad w_D = LS^{-1}.$$

The maximization of (3.10) over **T**, **U**, **V**, and **W** is done analogous to Kiers (1998a), and its description is omitted.

Chapter 4

Three-mode factor analysis by means of Candecomp/ Parafac

Abstract

A three-mode covariance matrix contains covariances of N observations (e.g., subject scores) on J variables for K different occasions or conditions. We model such an $JK \times JK$ covariance matrix as the sum of a (common) covariance matrix having Candecomp/Parafac form, and a diagonal matrix of unique variances. The Candecomp/Parafac form is a generalization of the two-mode case under the assumption of parallel factors. We estimate the unique variances by Minimum Rank Factor Analysis. The factors can be chosen oblique or orthogonal. Our approach yields a model that is easy to estimate and easy to interpret. Moreover, the unique variances, the factor covariance matrix and the communalities are guaranteed to be proper, a percentage of explained common variance can be obtained for each variable-condition combination, and the estimated model is rotationally unique under mild conditions. We apply our model to several datasets in the literature, and demonstrate our estimation procedure in a simulation study.

This chapter is based on: Stegeman, A. and Lam, T.T.T. (2014). Three-mode factor analysis by means of Candecomp/ Parafac. *Psychometrika*, 79(3), 426-443.

4.1 Introduction

Three-way data refer to data that can be arranged in a three-dimensional array or three-way array. Such data is found in many different contexts. For example: scores on various anxiety scales of a number of individuals in various situations; scores on various competences of a number of workers by several different assessors; scores on food quality indicators of a number of food products by several different judges; and fMRI brain scan measurements for different areas of the brain over a period of time for different individuals. The three sets of entities associated with such three-way data sets are called the three *modes* of the array.

In this chapter, we consider three-way data of N observations of J variables for K different occasions or conditions. We focus on three-mode factor analysis, i.e., a factor model for the $JK \times JK$ covariance matrix containing the covariances of all J variables and K conditions together. In section 4.1.1, we introduce the general framework of three-mode factor analysis. In section 4.1.2, we discuss existing models and methods for three-mode factor analysis, and introduce our novel model. In section 4.1.3, we discuss Minimum Rank Factor Analysis (MRFA) for two-way factor analysis. We use MRFA to estimate the unique variances in our method for three-mode factor analysis.

4.1.1 Three-mode factor analysis

Let $\mathbf{X}_k$ be an $N \times J$ matrix containing N observations of J variables for occasion k or under condition k, for $k = 1, \ldots, K$. We assume the columns of $\mathbf{X}_k$ have mean zero for all k. That is, offset terms that are constant across observations have been removed for each variable and condition. We suppose that in theory the data $\mathbf{X}_k$ can be written as the sum of a *common part* and a *unique part*: $\mathbf{X}_k = \mathbf{X}_k^{(\text{com})} + \mathbf{E}_k$, for $k = 1, \ldots, K$. The common part $\mathbf{X}_k^{(\text{com})}$ contains the

part of each variable under condition k that correlates with other variables in the data. The unique part $\mathbf{E}_k$ contains the part of each variable under condition k that does not correlate with other variables. The unique part of a variable may contain measurement error as well as a reliable part measuring a trait that is uncorrelated with any other variable. Both $\mathbf{X}_k^{(\text{com})}$ and $\mathbf{E}_k$ have mean-zero columns.

We look for a small number of R factors that best summarizes the common parts: $\mathbf{X}_k^{(\text{com})} \approx \mathbf{F}\,\mathbf{B}_k^T$, where the factors $\mathbf{F}$ $(N \times R)$ are the same for all k, but the loadings $\mathbf{B}_k$ $(J \times R)$ may be different. For all $\mathbf{X}_k$ together, and perfect fit, we have the following factor model:

$$\mathbf{X}_{(N\times JK)} = \mathbf{F} \begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_K \end{bmatrix}^T + \mathbf{E}_{(N\times JK)}\,, \tag{4.1}$$

where $\mathbf{X}_{(N\times JK)} = [\mathbf{X}_1 \ldots \mathbf{X}_K]$ and $\mathbf{E}_{(N\times JK)} = [\mathbf{E}_1 \ldots \mathbf{E}_K]$. The covariance model corresponding to (4.1) is

$$\mathbf{\Sigma} \;=\; \begin{bmatrix} \mathbf{\Sigma}_{11} & \cdots & \mathbf{\Sigma}_{1K} \\ \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{K1} & \cdots & \mathbf{\Sigma}_{KK} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_K \end{bmatrix} \mathbf{\Phi} \begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_K \end{bmatrix}^T + \mathbf{U}\,, \tag{4.2}$$

where $\mathbf{\Sigma} = N^{-1}\mathbf{X}_{(N\times JK)}^T\mathbf{X}_{(N\times JK)}$ is the data covariance matrix, $\mathbf{\Sigma}_{kl} = N^{-1}\mathbf{X}_k^T\mathbf{X}_l$ contains the covariances between the J variables for conditions k and l, the factor covariance matrix is $\mathbf{\Phi} = N^{-1}\mathbf{F}^T\mathbf{F}$, and $\mathbf{U} = N^{-1}\mathbf{E}_{(N\times JK)}^T\mathbf{E}_{(N\times JK)}$ is the diagonal matrix of unique variances. Note that $\mathbf{\Sigma}$ and $\mathbf{U}$ have size $JK \times JK$, that $\mathbf{\Sigma}_{kl}$ has size $J \times J$, and that $\mathbf{\Sigma}_{kl} = \mathbf{\Sigma}_{lk}^T$.

The factors $\mathbf{F}$ are usually scaled such that they have variance 1, which makes $\mathbf{\Phi}$ the factor correlation matrix. If the factors are chosen uncorrelated (also

called *orthogonal*), then $\mathbf{\Phi} = \mathbf{I}_R$. Otherwise, the factors are called *oblique*. The diagonal entries of $\mathbf{\Sigma} - \mathbf{U}$ are the variances of the common parts of the variables for each condition, and are called *communalities* or common variances. The diagonal entries of $\mathbf{B}_{(\text{all})}\mathbf{\Phi}\mathbf{B}_{(\text{all})}^T$ are called the estimated common variances, where $\mathbf{B}_{(\text{all})} = [\mathbf{B}_1^T \ \ldots \ \mathbf{B}_K^T]^T$. The diagonal entries of $\mathbf{U}$ are called the unique variances.

A probabilistic version of the three-mode factor model (4.1) is

$$\begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_K \end{pmatrix} = \begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_K \end{bmatrix} \mathbf{f} + \begin{pmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_K \end{pmatrix}, \tag{4.3}$$

where $\mathbf{x}_k$ is the $J \times 1$ random vector corresponding to the J variables under condition k, the $R \times 1$ random vector $\mathbf{f}$ contains the factors, and the $J \times 1$ random vector $\mathbf{e}_k$ corresponds to the unique parts of the variables under condition k. It is assumed that $\mathbf{x}_k$, $\mathbf{f}$, and $\mathbf{e}_k$ have zero expectation for all k, that $\mathbf{f}$ and $\mathbf{e}_k$ are uncorrelated for all k, and that $\mathbf{e}_k$ and $\mathbf{e}_l$ are uncorrelated for $k \neq l$. Under these assumptions, the covariance model corresponding to (4.3) is equal to (4.2). The covariances between the variables and factors are given by $\mathbf{B}_{(\text{all})}\mathbf{\Phi}$. The correlations between variables and factors are used to interpret the factors when the fit of the factor model is high.

Some examples of data for which three-mode factor analysis models may be useful, are the following:

- A depression scale of J items filled in by N persons on K time points. The first measurement may be pre-treatment, while the consecutive measurements include the effect of the treatment. A three-mode factor model shows the loadings $\mathbf{B}_k$ for each measurement k on the same factors $\mathbf{F}$.

- Multitrait-multimethod data, where J traits are measured for N persons, using K different methods. Here, the loadings $\mathbf{B}_k$ are method-specific. See below for a discussion, and section 4.3.1 for an example.

- A belief in a just world scale of J items filled in by N persons, where each item is both formulated as "justice for yourself" and "justice for others". Hence, we have $K = 2$ conditions. See section 4.3.2 for a worked example.

So far, we have presented a general form of three-mode factor analysis. Various specific forms for the loadings $\mathbf{B}_k$ have been proposed in the literature. In the next subsection, we will give a short overview, and introduce our own model as an alternative. While existing models may suffer from estimation problems (e.g., convergence problems or non-admissable solutions), identification problems, interpretation difficulties, or lack of parsimony, our model does not have these shortcomings, as we will see in section 4.1.2.

4.1.2 Models for three-mode factor analysis

An important application of three-mode factor analysis is on multitrait-multimethod data, where the J variables measure one or several personality traits of N individuals, using K different methods. The covariance matrix $\mathbf{\Sigma}$ can be used to study trait validity across the different measurement methods (Campbell & Fiske, 1959). In this context, it has been proposed to assign each factor to a specific trait or to a specific method (e.g. Widaman, 1985). For example, if $J = 3$ variables measure three traits and $K = 2$ methods are used, then $R = 5$

factors may be included with loadings

$$\mathbf{B}_{(\text{all})} = \left[\begin{array}{c} \mathbf{B}_1 \\ \hline \mathbf{B}_2 \end{array}\right] = \left[\begin{array}{ccccc} * & 0 & 0 & * & 0 \\ 0 & * & 0 & * & 0 \\ 0 & 0 & * & * & 0 \\ \hline * & 0 & 0 & 0 & * \\ 0 & * & 0 & 0 & * \\ 0 & 0 & * & 0 & * \end{array}\right],$$

where $*$ denotes an arbitrary nonzero entry. Hence, the first three factors correspond to the three traits, and the last two factors correspond to the two methods. Additionally, it may be required that the trait factors are uncorrelated with the two method factors. This can be done by constraining the corresponding covariances in $\mathbf{\Phi}$ to zero. In that case, the estimated common variances on the diagonal of $\mathbf{B}_{(\text{all})}\,\mathbf{\Phi}\,\mathbf{B}_{(\text{all})}^T$ can be written as the sum of a part due to the trait factors and a part due to the method factors. That is, it holds that

$$\mathbf{B}_{(\text{all})}\,\mathbf{\Phi}\,\mathbf{B}_{(\text{all})}^T = \mathbf{B}_{(\text{trait})}\,\mathbf{\Phi}_1\,\mathbf{B}_{(\text{trait})}^T + \mathbf{B}_{(\text{meth})}\,\mathbf{\Phi}_2\,\mathbf{B}_{(\text{meth})}^T, \tag{4.4}$$

where $\mathbf{B}_{(\text{trait})}$ and $\mathbf{B}_{(\text{meth})}$ contain the columns of $\mathbf{B}_{(\text{all})}$ corresponding to trait and method factors, respectively, and $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are the covariance matrices of the trait and method factors, respectively.

These type of models are confirmatory factor analysis models and can be estimated by Maximum Likelihood (MLFA) (Jöreskog, 1970, 1971b). For a detailed overview of this approach, and the related method of covariance component analysis, we refer to Wothke (1996). Identification results for factor models of this type can be found in Millsap (1992). Some problems may occur when fitting confirmatory factor analysis models for multitrait-multimethod data. Both convergence problems and improper solutions (e.g., where $\mathbf{\Phi}$ or $\mathbf{\Sigma} - \mathbf{U}$ is not

a covariance matrix) are well known (Kiers, Takane, & Ten Berge, 1996). To overcome these problems, Kiers et al. (1996) proposed to fit constrained component models on the data instead. Here, the unique parts are treated as errors or noise, and the factor model (4.1) with constrained $\mathbf{B}_{(\text{all})}$ and $\mathbf{\Phi}$ is fitted directly to the data instead of fitting covariances. However, this alternative method does not overcome the problems in all cases.

Eid (2000) has proposed a class of confirmatory factor models for multitrait-multimethod data based on classical psychometric test theory. In such models, the number of method factors is one less than the number of methods. As a result, one method is used as a comparison standard. Eid (2000) shows that his models are globally identified. Eid et al. (2008) discuss how to choose an appropriate confirmatory factor model for different types of methods in multitrait-multimethod data. Although the approach based on Eid (2000) solves the identification problems and seems to solve the convergence problems of confirmatory factor models, it does not always produce a proper factor covariance matrix. For example, the factor covariance matrix in table 4 of Eid et al. (2008) has many negative eigenvalues. Also, the fit of the model depends on which method is chosen as comparison standard, a choice which may not be obvious from a substantive point of view.

A different approach to three-mode factor analysis is based on the three-mode component model by Tucker (1966). In this model, each mode of the data has its own components, and their interaction strengths are given by numbers g_{rpq} of the so called *core array*. For our three-way data, suppose we have R components for the N observations, P components for the J variables, and Q components for the K conditions. The Tucker3 model can then be written in

the form of (4.1) as

$$\mathbf{X}_{(N\times JK)} = \mathbf{F}\,\mathbf{G}\,(\mathbf{C}\otimes\mathbf{B})^T + \mathbf{E}_{(N\times JK)}\,, \tag{4.5}$$

where $\mathbf{C}$ is an $K \times Q$ matrix containing the Q method components as columns, $\mathbf{B}$ is a $J \times P$ matrix containing the P variable components as columns, $\mathbf{C}\otimes\mathbf{B}$ is the right direct or Kronecker product of $\mathbf{C}$ and $\mathbf{B}$, and $\mathbf{G}$ is the $R \times PQ$ matrix of interaction strengths, with

$$\mathbf{G} = \left[\begin{array}{ccc|c|ccc} g_{111} & \cdots & g_{1P1} & \cdots & g_{11Q} & \cdots & g_{1PQ} \\ \vdots & & \vdots & & \vdots & & \vdots \\ g_{R11} & \cdots & g_{RP1} & \cdots & g_{R1Q} & \cdots & g_{RPQ} \end{array}\right].$$

It is well known that the three-mode Tucker model is not unique. All of $\mathbf{F}$, $\mathbf{C}$, and $\mathbf{B}$ can be rotated, with inverse transformations applied to the interaction strengths in $\mathbf{G}$, without affecting the model part $\mathbf{FG}(\mathbf{C}\otimes\mathbf{B})^T$. The covariance model corresponding to (4.5) is

$$\mathbf{\Sigma} = (\mathbf{C}\otimes\mathbf{B})\,\mathbf{G}^T\,\mathbf{\Phi}\,\mathbf{G}\,(\mathbf{C}\otimes\mathbf{B})^T + \mathbf{U}\,. \tag{4.6}$$

The associated probabilistic model (4.3) has been introduced by Bloxom (1968) and is further analyzed by Bentler and Lee (1978, 1979). For simplicity, one may consider $\mathbf{\Psi} = \mathbf{G}^T\mathbf{\Phi}\mathbf{G}$ as factor covariance matrix. The case of diagonal $\mathbf{\Psi}$ can be rewritten as $\mathbf{\Sigma} = (\mathbf{C}\mathbf{C}^T \otimes \mathbf{B}\mathbf{B}^T) + \mathbf{U}$ and is known as the direct product model; see Browne (1984) or Wothke (1996). In this model, $\mathbf{C}\mathbf{C}^T$ and $\mathbf{B}\mathbf{B}^T$ may be seen as covariance matrices corresponding to methods and variables, respectively. The estimated common variances on the diagonal of $\mathbf{C}\mathbf{C}^T\otimes\mathbf{B}\mathbf{B}^T$ are obtained by multiplying the corresponding variances in $\mathbf{C}\mathbf{C}^T$ (for the method) and $\mathbf{B}\mathbf{B}^T$ (for the variable). In this sense, the direct product model is multiplicative, whereas confirmatory factor analysis models are additive; see (4.4).

Fitting the covariance model (4.6) can be done by directly fitting the component model (4.5) to the data. Alternating least squares algorithms minimizing the sum-of-squares of $\mathbf{E}_{(N\times JK)}$ can be found in Kroonenberg and De Leeuw (1980) and Kiers, Kroonenberg, and Ten Berge (1992). Contrary to the confirmatory factor analysis approach, convergence problems do not often occur. One may also fit the covariance model (4.6) by an algorithm for nonlinear optimization. For example, Bentler and Lee (1978, 1979) propose to use a Gauss-Newton algorithm. For an overview of three-mode component and factor models based on Tucker (1966), we refer to Kroonenberg and Oort (2003). For an accessible introduction to three-mode component analysis, see Kiers and Van Mechelen (2001).

As an alternative to confirmatory factor analysis and the three-mode component model of Tucker (1966), we propose a special case of the latter. Namely, the three-mode component model known as Candecomp (Carroll & Chang, 1970) or Parafac (Harshman, 1970), which was originally introduced in mathematics (Hitchcock, 1927a, 1927b). The Candecomp/Parafac model is a special case of (4.5) in which $R = P = Q$ and $g_{rrr} = 1$ and $g_{rpq} = 0$ otherwise. Hence, we have an equal number of components in each mode, and there are no interactions between different component numbers. Analogous to (4.5), the Candecomp/Parafac model is given by

$$\mathbf{X}_{(N\times JK)} = \mathbf{F}\,(\mathbf{C}\odot\mathbf{B})^T + \mathbf{E}_{(N\times JK)}\,, \tag{4.7}$$

where $\mathbf{C}\odot\mathbf{B} = [\mathbf{c}_1\otimes\mathbf{b}_1|\,\ldots\,|\mathbf{c}_R\otimes\mathbf{b}_R]$ is the (column-wise) Khatri-Rao product, and contains Kronecker products between the corresponding pairs of columns of $\mathbf{C}$ and $\mathbf{B}$. In the general three-mode factor model (4.1), the Candecomp/Parafac model corresponds to the case $\mathbf{B}_k = \mathbf{B}\,\mathbf{C}_k$, where $\mathbf{C}_k$ is the diagonal matrix with row k of $\mathbf{C}$ as its diagonal. In the Candecomp/Parafac model part, the matrices

$\mathbf{F}$, $\mathbf{C}$, and $\mathbf{B}$ are unique up to scaling and a simultaneous column permutation under e.g. the condition (Kruskal, 1977)

$$k_{\mathbf{F}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2\,R + 2\,, \tag{4.8}$$

where $k_{\mathbf{Y}}$ denotes the k-rank of a matrix $\mathbf{Y}$. The latter is defined as the largest number x such that every subset of x columns of $\mathbf{Y}$ is linearly independent. Other, more relaxed uniqueness conditions exist. However, (4.8) is easy to check and satisfies our current needs.

The covariance model corresponding to (4.7) is given by

$$\mathbf{\Sigma} = (\mathbf{C} \odot \mathbf{B})\,\mathbf{\Phi}\,(\mathbf{C} \odot \mathbf{B})^T + \mathbf{U}\,. \tag{4.9}$$

Also in the common covariance part of (4.9), matrices $\mathbf{C}$, $\mathbf{B}$, and $\mathbf{\Phi}$ are unique up to scaling and permutation when (4.8) holds. For orthogonal factors, the covariance matrix of the common part can be written as

$$(\mathbf{C} \odot \mathbf{B})(\mathbf{C} \odot \mathbf{B})^T = \sum_{r=1}^{R} (\mathbf{c}_r \otimes \mathbf{b}_r)(\mathbf{c}_r \otimes \mathbf{b}_r)^T = \sum_{r=1}^{R} (\mathbf{c}_r \mathbf{c}_r^T \otimes \mathbf{b}_r \mathbf{b}_r^T)\,,$$

where $\mathbf{c}_r$ and $\mathbf{b}_r$ denote column r of $\mathbf{C}$ and $\mathbf{B}$, respectively. The estimated common variance for condition k and variable j can be expressed as $\sum_{r=1}^{R} c_{kr}^2 b_{jr}^2$. This shows that the estimated common variances in the Candecomp/Parafac covariance model (4.9) are obtained by both multiplications of method and variable coefficients, and additions over the number of factors.

Our estimation procedure for the Candecomp/Parafac covariance model first uses Minimum Rank Factor Analysis (see section 4.1.3) to compute $\mathbf{U}$, which guarantees proper communalities on the diagonal of $\mathbf{\Sigma} - \mathbf{U}$. This is not the case for both confirmatory factor analysis models and the three-mode factor model based on Tucker (1966) (in the way it has been estimated so far). Next, we estimate $\mathbf{C}$, $\mathbf{B}$, and $\mathbf{\Phi}$ by the alternating least squares algorithm of Harshman

(1970). We are able to compute the total percentage of explained common variance, as well as for each variable-condition combination separately. Also, the factor covariance matrix $\mathbf{\Phi}$ is guaranteed to be proper (i.e., have nonnegative eigenvalues). Our algorithm does not often suffer from convergence problems. Compared to the Tucker model of covariances (4.6), we only need to determine the number R of components instead of R, P, and Q. But a special case of Candecomp/Parafac also enables a choice of $R = P$ and $Q = 1$. Also, our Candecomp/Parafac covariance model (4.9) is unique under mild conditions, while the Tucker-based covariance model (4.6) is not. Analogous to two-way factor analysis methods, the common part of the covariance model (4.6) can be rotated (orthogonally or obliquely) to an approximately sparse matrix $\mathbf{G}$ of interaction strengths (Kiers, 1998b, 1998a; Tendeiro, Ten Berge, & Kiers, 2009). Hence, using the covariance model (4.6) involves a choice of rotation method.

Whether uniqueness in three-way factor analysis is desirable or not is not agreed upon. For example, as an anonymous reviewer stated, in psychology researchers are used to rotating two-way factor solutions and often consider the rotational freedom an advantage from an interpretational point of view. However, as we will illustrate, our unique Candecomp/Parafac covariance model is also easy to interpret. Still, when the rotational freedom and interactions between different components of the Tucker-based covariance model are desired, it can also be incorporated in our estimation scheme. In this paper, however, we focus on the Candecomp/Parafac covariance model. A discussion on model selection and the comparison between Candecomp/Parafac and the model of Tucker (1966) can be found in section 4.5.

Three-mode component models such as Tucker (1966) and Candecomp/Parafac and their four-mode and higher-mode extensions are applied in the social sciences (Kroonenberg, 2008), chemometrics (Smilde, Bro, & Geladi, 2004), independent

component analysis (Comon & De Lathauwer, 2010; De Lathauwer, 2010), and data mining in general. For an overview of applications, see Kolda and Bader (2009) or Acar and Yener (2009).

4.1.3 Minimum Rank Factor Analysis

For later use, we discuss in some detail the Minimum Rank Factor Analysis (MRFA) method for two-mode factor analysis. Here, the covariance model is $\mathbf{\Sigma} \approx \mathbf{B\Phi B}^T + \mathbf{U}$, where $\mathbf{\Sigma}$ is the $J \times J$ data covariance matrix, $\mathbf{B}$ is the $J \times R$ loadings matrix, and $\mathbf{U}$ is the $J \times J$ diagonal matrix of unique variances. MRFA is used to estimate $(\mathbf{B}, \mathbf{\Phi}, \mathbf{U})$; see Ten Berge and Kiers (1991). The MRFA algorithm computes the unique variances $\mathbf{U}$ such that $\mathbf{U}$ is nonnegative, $\mathbf{\Sigma} - \mathbf{U}$ is a covariance matrix, and the unexplained common variance in $\mathbf{\Sigma} - \mathbf{U} \approx \mathbf{B\Phi B}^T$ is minimized. The matrix $\mathbf{\Sigma} - \mathbf{U}$ is a covariance matrix if it is equal to $\mathbf{H}^T\mathbf{H}$ for some matrix $\mathbf{H}$. This is equivalent to $\mathbf{\Sigma} - \mathbf{U}$ having nonnegative eigenvalues. When using MINRES or MLFA, it may happen that $\mathbf{\Sigma} - \mathbf{U}$ is not a covariance matrix. The best approximation $\mathbf{B\Phi B}^T$ is obtained from the R largest eigenvalues and associated eigenvectors of $\mathbf{\Sigma} - \mathbf{U}$, and the minimum unexplained common variance in $\mathbf{\Sigma} - \mathbf{U} \approx \mathbf{B\Phi B}^T$ is equal to the sum of the $J - R$ smallest eigenvalues of $\mathbf{\Sigma} - \mathbf{U}$; see Eckart and Young (1936).

The advantage of MRFA is that we have proper communalities and we can compute the percentage of explained common variance as

$$100 \cdot \frac{\text{trace}(\mathbf{B}\,\mathbf{\Phi}\,\mathbf{B}^T)}{\text{trace}(\mathbf{\Sigma} - \mathbf{U})}\,, \tag{4.10}$$

where trace$(\cdot)$ is defined as the sum of the diagonal entries of a matrix, which is equal to the sum of the eigenvalues of the matrix. The numerator of (4.10) equals the sum of the estimated common variances. The denominator equals the sum of the communalities (common variances), under the condition that

the eigenvalues of $\mathbf{\Sigma} - \mathbf{U}$ are nonnegative. To sum up, for a fixed number of R factors, MRFA minimizes the amount of common variance left unexplained under the constraint of proper communalities. A detailed comparison between MRFA and other factor analysis methods can be found in Sŏcan (2003).

Finally, we give an outline of the remaining part of the chapter. In section 4.2, we present our estimation procedure for the Candecomp/Parafac covariance model (4.9). In section 4.3, we apply our Candecomp/Parafac method to several datasets in the literature, and compare the results with other methods. In section 4.4, we assess the performance of our estimation procedure in a simulation study. Section 4.5 contains a discussion of our findings.

4.2 Three-mode factor analysis by means of Candecomp/Parafac

Here, we present our estimation procedure for the Candecomp/Parafac covariance model. After having computed the data covariance matrix $\mathbf{\Sigma}$, the steps of our estimation procedure for the Candecomp/Parafac covariance model (4.9) are as follows.

1. Use the MRFA algorithm of Ten Berge and Kiers (1991) to estimate $\mathbf{U}$. This implies that $\mathbf{U}$ is nonnegative, $\mathbf{\Sigma} - \mathbf{U}$ is a covariance matrix, and the trace of $\mathbf{\Sigma} - \mathbf{L} - \mathbf{U}$ is minimal, where $\mathbf{L}$ is a best rank-R approximation of $\mathbf{\Sigma} - \mathbf{U}$.

2. Compute the eigendecomposition $\mathbf{\Sigma} - \mathbf{U} = \mathbf{V}\,\mathbf{S}\,\mathbf{V}^T$, with $\mathbf{V}$ having orthonormal columns, and $\mathbf{S}$ the diagonal matrix containing the eigenvalues in decreasing order. This is also the singular value decomposition of $\mathbf{\Sigma} - \mathbf{U}$. Let $\mathbf{P} = \mathbf{V}\mathbf{S}^{1/2}$, which implies $\mathbf{\Sigma} - \mathbf{U} = \mathbf{P}\mathbf{P}^T$. If $\mathbf{P}_R$ contains the first R

columns of $\mathbf{P}$, then $\mathbf{L} = \mathbf{P}_R\mathbf{P}_R^T$ is a best rank-R approximation of $\mathbf{\Sigma} - \mathbf{U}$; see Eckart and Young (1936). In the next step, we approximate $\mathbf{P}$ by a Candecomp/Parafac decomposition with R components.

3. Fit the Candecomp/Parafac model as $\mathbf{P} \approx (\mathbf{C} \odot \mathbf{B})\,\mathbf{T}^T$ by using the alternating least squares algorithm of Harshman (1970). It is recommended to do e.g. ten runs of the algorithm with random starting values, and to keep the best run as Candecomp/Parafac solution. Matrix $\mathbf{P}$ is a matricized $K \times J \times JK$ array, and $\mathbf{T}$ is a $JK \times R$ matrix. The columns of $\mathbf{T}$ are scaled such that they have length 1. For orthogonal factors, restrict the columns of $\mathbf{T}$ to be orthogonal. We obtain $\mathbf{\Sigma}-\mathbf{U} \approx (\mathbf{C}\odot\mathbf{B})\,\mathbf{T}^T\mathbf{T}\,(\mathbf{C}\odot\mathbf{B})^T$. We evaluate the Candecomp/Parafac fit as

$$100 - 100 \cdot \frac{\text{ssq}(\mathbf{P} - (\mathbf{C} \odot \mathbf{B})\,\mathbf{T}^T)}{\text{ssq}(\mathbf{P})}\,, \tag{4.11}$$

which is the percentage of the sum-of-squares of $\mathbf{P}$ that is fitted by $(\mathbf{C} \odot \mathbf{B})\,\mathbf{T}^T$. In the alternating least squares algorithm, $(\mathbf{C} \odot \mathbf{B})\,\mathbf{T}^T$ is the regression of $\mathbf{P}$ on $(\mathbf{C} \odot \mathbf{B})$. Since the regression and the residual are orthogonal, it follows that (4.11) is equal to $100 \cdot \text{ssq}((\mathbf{C} \odot \mathbf{B})\,\mathbf{T}^T)/\text{ssq}(\mathbf{P})$.

4. For oblique factors, let $\mathbf{\Phi} = \mathbf{T}^T\mathbf{T}$. For orthogonal factors, we have $\mathbf{\Phi} = \mathbf{T}^T\mathbf{T} = \mathbf{I}_R$.

Since we were not able to construct an algorithm for the simultaneous estimation of $\mathbf{U}$ and $\mathbf{C}, \mathbf{B}, \mathbf{\Phi}$ under the restriction that $\mathbf{\Sigma} - \mathbf{U}$ is a covariance matrix, we instead estimate $\mathbf{U}$ and $\mathbf{C}, \mathbf{B}, \mathbf{\Phi}$ sequentially. First, we estimate $\mathbf{U}$ by MRFA based on a rank-R factor model for $\mathbf{\Sigma} - \mathbf{U}$. For fitting $(\mathbf{C} \odot \mathbf{B})\,\mathbf{\Phi}\,(\mathbf{C} \odot \mathbf{B})^T$ to $\mathbf{\Sigma} - \mathbf{U}$, we have considered two options. Either we approximate the best rank-R approximation of $\mathbf{\Sigma} - \mathbf{U}$ by the Candecomp/Parafac covariance part as

$\mathbf{P}_R \approx (\mathbf{C} \odot \mathbf{B})\,\mathbf{T}_R^T$, where $\mathbf{T}_R$ is an $R \times R$ matrix, or we use steps 2 and 3 above, in which the Candecomp/Parafac covariance part is fitted on the full $\mathbf{P}$. Using the approximation of $\mathbf{P}_R$ has as advantage that it is consistent with the estimation of $\mathbf{U}$, which assumes a rank-R factor model. Moreover, we observed that the fit percentage of $\mathbf{P}_R \approx (\mathbf{C} \odot \mathbf{B})\,\mathbf{T}_R^T$ is usually above 99 percent for datasets in the literature, which reduces the need for a simultaneous estimation procedure. However, when using the approximation of $\mathbf{P}_R$ it cannot be guaranteed that the estimated common variance is at most equal to the communality for each combination of variable and condition. Therefore, we have chosen to use the approximation of $\mathbf{P}$ instead, for which this is guaranteed (see further below). Conceptually, this introduces an inconsistency between the two estimation steps. However, the performance of both variants of the estimation procedure is very similar in our simulation study.

In the obtained matrices $\mathbf{C}$ and $\mathbf{B}$, there is still a scaling indeterminacy for each pair of columns (i.e., $(\lambda\,\mathbf{c}_r) \otimes (\lambda^{-1}\,\mathbf{b}_r) = \mathbf{c}_r \otimes \mathbf{b}_r$). This can be fixed by rescaling the columns of $\mathbf{B}$ to length 1. Note that step 4 guarantees a factor covariance matrix $\mathbf{\Phi}$ that is proper, i.e., has nonnegative eigenvalues.

The Candecomp/Parafac matrices $\mathbf{C}, \mathbf{B}, \mathbf{T}$ are unique up to permutation and scaling under mild conditions; e.g. (4.8). When both $(\mathbf{C} \odot \mathbf{B})$ and $\mathbf{T}$ have rank R, then also the Candecomp/Parafac covariance model $(\mathbf{C} \odot \mathbf{B})\,\mathbf{\Phi}\,(\mathbf{C} \odot \mathbf{B})^T$ is unique up to permutation and scaling. Note that $(\mathbf{C} \odot \mathbf{B})$ must have rank R for Candecomp/Parafac uniqueness (Liu & Sidiropoulos, 2001; Stegeman & Sidiropoulos, 2007). When desired, the Tucker-based covariance model can replace our Candecomp/Parafac covariance model in step 3 above. In that case, $\mathbf{P} \approx (\mathbf{C} \otimes \mathbf{B})\,\mathbf{G}^T\,\mathbf{T}^T$ is estimated. However, this is outside the scope of this chapter.

For orthogonal factors, it has been proven that a best-fitting Candecomp/Parafac model $\mathbf{P} \approx (\mathbf{C} \odot \mathbf{B})\,\mathbf{T}^T$ always exists (Krijnen et al., 2008). For oblique factors, this is not necessarily true, and one may encounter convergence problems and uninterpretable solutions (Harshman & Lundy, 1984; Kruskal et al., 1989; Paatero, 2000; Harshman, 2004; Stegeman, 2006, 2007, 2008, 2009b; Krijnen et al., 2008; De Silva & Lim, 2008). However, this problem has not occurred when analyzing reallife datasets such as those in section 4.3. There are no general results stating conditions on a three-way array under which a best-fitting Candecomp/Parafac model exists. This is still an open problem (De Silva & Lim, 2008; Stegeman, 2006, 2007, 2008).

As goodness-of-fit measure we wish to use the percentage of explained common variance. The percentage of explained common variance in our Candecomp/Parafac covariance model is given by (4.11), which can be written as

$$100 \cdot \frac{\mathrm{trace}((\mathbf{C} \odot \mathbf{B})\,\mathbf{T}^T\mathbf{T}\,(\mathbf{C} \odot \mathbf{B})^T)}{\mathrm{trace}(\mathbf{\Sigma} - \mathbf{U})}\,. \tag{4.12}$$

If the factors are chosen orthogonal, then we may also obtain a percentage of explained common variance due to each factor. Namely, $\mathbf{T}^T\mathbf{T} = \mathbf{I}_R$ in (4.12), and we can define the explained common variance due to factor r as the sum-of-squares of column r of $(\mathbf{C} \odot \mathbf{B})$. This column is $\mathbf{c}_r \otimes \mathbf{b}_r$ and has sum-of-squares equal to $(\mathbf{c}_r^T\mathbf{c}_r)(\mathbf{b}_r^T\mathbf{b}_r)$. Hence, we have

$$100 \cdot \frac{\mathrm{trace}((\mathbf{C} \odot \mathbf{B})(\mathbf{C} \odot \mathbf{B})^T)}{\mathrm{trace}(\mathbf{\Sigma} - \mathbf{U})} \;=\; \sum_{r=1}^{R} \left(100 \cdot \frac{(\mathbf{c}_r^T\mathbf{c}_r)(\mathbf{b}_r^T\mathbf{b}_r)}{\mathrm{trace}(\mathbf{\Sigma} - \mathbf{U})} \right), \tag{4.13}$$

where the summands in the right-hand side express the percentage of explained common variance due to each factor.

To obtain a percentage of explained common variance for each variable-condition combination separately, we proceed as follows. Contrary to the two-mode case of MRFA, it may happen that some diagonal entry of $(\mathbf{C}\odot\mathbf{B})\,\mathbf{T}^T\mathbf{T}\,(\mathbf{C}\odot$

$\mathbf{B})^T$ is larger than the corresponding communality on the diagonal of $\mathbf{\Sigma}-\mathbf{U}$. Because of this, we formulate the explained common variance per variable-condition combination analogous to (4.11) rather than to (4.12). A particular variable-condition corresponds to a row of $\mathbf{P} - (\mathbf{C} \odot \mathbf{B})\,\mathbf{T}^T$. Let row m of this matrix be denoted as $\mathbf{q}_m^T$. Then we define the corresponding percentage of explained common variance as

$$100 - 100 \cdot \frac{\text{ssq}(\mathbf{q}_m^T)}{(\mathbf{\Sigma} - \mathbf{U})_{mm}},$$

where $(\mathbf{\Sigma} - \mathbf{U})_{mm}$ is the corresponding communality.

In the Candecomp/Parafac covariance model (4.9) we have R components in the individuals mode, in the condition mode, and in the variable mode. Using a constrained Candecomp/Parafac model, it is possible to have different numbers of components in different modes. For example, instead of the Candecomp/Parafac model $(\mathbf{C}, \mathbf{B}, \mathbf{T})$ with $R = 2$, we can have $(\mathbf{c\Psi}, \mathbf{B}, \mathbf{T})$, with $\mathbf{\Psi} = [1 \quad 1]$ being fixed. Hence, we have one component $\mathbf{c}$ in the condition mode, which interacts with both components in the individuals and variables modes. This type of Candecomp/Parafac model is known as Paralind (Bro, Harshman, Sidiropoulos, & Lundy, 2009) or Confac (De Almeida, Favier, & Mota, 2008b, 2008a), and can be fitted by an alternating least squares algorithm. Uniqueness conditions for Paralind models are proven in Stegeman and Almeida (2009) and Stegeman and Lam (2012). The general Paralind model has form $(\mathbf{C\Psi}, \mathbf{B\Phi}, \mathbf{T\Omega})$, where $\mathbf{C}, \mathbf{B}, \mathbf{T}$ have linearly independent columns, and $\mathbf{\Psi}, \mathbf{\Phi}, \mathbf{\Omega}$ are fixed constraint matrices.

4.3 Application to datasets in the literature

Here, we apply our three-mode factor model (4.9) to datasets in the literature. We use the estimation procedure outlined in section 4.2. In section 4.3.1, we

consider a multitrait-multimethod (MTMM) dataset with four personality traits and two methods. Using a three-mode factor model (4.6) based on the Tucker (1966) three-mode component model, this dataset was analyzed by Bentler and Lee (1978). We show that our Candecomp/Parafac approach can incorporate the model of Bentler and Lee (1978), but has lower unique variances and a proper covariance matrix $\mathbf{\Sigma} - \mathbf{U}$. The latter is not the case in Bentler and Lee (1978). In section 4.3.2, we consider a belief in a just world scale with eight items, where each item is asked from two perspectives: a just world for oneself and a just world for others. The data is part of a recent study among online American respondents.

4.3.1 MTMM data from Bentler and Lee (1978)

In this dataset, $J = 4$ personality traits are measured with $K = 2$ methods. The traits are ambition, attractiveness, leadership, and extraversion. The methods are self-report and peer-report. The number of individuals in the dataset is $N = 72$. The correlation matrix $\mathbf{\Sigma}$ reported in Bentler and Lee (1978) is given in Table 4.1 below.

In Bentler and Lee (1978), the three-mode factor model (4.6) is used with $Q = 1$ method component, $P = 2$ components for the variables, and $R = 2$ orthogonal factors. The covariance model (4.6) is simplified to $\mathbf{G}^T\mathbf{\Phi}\mathbf{G} = \mathbf{I}_2$. As estimation procedure for (4.6), Bentler and Lee (1978) use a Gauss-Newton algorithm. Their results are:

$$\mathbf{C} = \begin{pmatrix} 1 \\ 0.85 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 0.74 & 0 \\ 0.30 & 0.36 \\ 0.41 & 0.71 \\ 0.27 & 0.80 \end{pmatrix}, \tag{4.14}$$

Table 4.1: Correlations of personality variables ambition (Am), attractiveness (At), leadership (Le), and extraversion (Ex), measured by self-report and peer-report.

	self-report				peer-report			
variable	Am	At	Le	Ex	Am	At	Le	Ex
Am	1	0.223	0.337	0.223	0.402	0.035	0.160	0.093
At	0.223	1	0.418	0.290	0.070	0.442	0.196	0.180
Le	0.337	0.418	1	0.693	0.226	0.251	0.603	0.451
Ex	0.223	0.290	0.693	1	0.210	0.219	0.639	0.645
Am	0.402	0.070	0.226	0.210	1	0.233	0.379	0.269
At	0.035	0.442	0.251	0.219	0.233	1	0.314	0.283
Le	0.160	0.196	0.603	0.639	0.379	0.314	1	0.582
Ex	0.093	0.180	0.451	0.645	0.269	0.283	0.582	1

Note: Data taken from Bentler and Lee (1978)

and $\text{diag}(\mathbf{U}) = (0.63 \quad 0.71 \quad 0.52 \quad 0.47 \quad 0.69 \quad 0.74 \quad 0.58 \quad 0.63)$. The non-uniqueness of $\mathbf{C}$ and $\mathbf{B}$ has been fixed by setting the first entry of $\mathbf{C}$ equal to one, and rotating $\mathbf{B}$ such that its entry (1,2) is zero. The eigenvalues of $\mathbf{\Sigma} - \mathbf{U}$ are: -0.31, -0.29, -0.27, -0.13, 0.25, 0.43, 0.52, 2.83. The negative eigenvalues make $\text{tr}(\mathbf{\Sigma} - \mathbf{U})$ useless as a measure of total common variance. The model has 8 parameters (not counting the unique variances) and the sum-of-squares of $\mathbf{\Sigma} - \mathbf{U} - (\mathbf{C} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{B})^T$ equals 0.67.

The interpretation of (4.14) could be as follows. The same factors underly the variables for self-report and for peer-report. In the peer-report condition, the factors are a bit less pronounced (weight 0.85 opposed to weight 1 for self-report). The first factor can be interpreted as ambition (variable 1), and the

second factor as leadership and extraversion (variables 3 and 4). The two factors are uncorrelated. Variable 2 (attractiveness) is not a good indicator of either of the two factor dimensions.

Next, we estimate the Candecomp/Parafac covariance model (4.9) with $R = 2$ oblique factors, using the estimation procedure in section 4.2. That is, we fit $\mathbf{P} \approx (\mathbf{C} \odot \mathbf{B})\,\mathbf{T}^T$ with $\mathbf{P}$ such that $\mathbf{\Sigma} - \mathbf{U} = \mathbf{P}\mathbf{P}^T$. The results are:

$$\mathbf{C} = \begin{pmatrix} 1.56 & 1.29 \\ 1.37 & 0.56 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0.39 & 0.90 \\ 0.36 & 0.25 \\ 0.62 & 0.32 \\ 0.58 & 0.15 \end{pmatrix}, \quad \mathbf{\Phi} = \begin{pmatrix} 1 & -0.53 \\ -0.53 & 1 \end{pmatrix}, \tag{4.15}$$

and $\text{diag}(\mathbf{U}) = (0 \quad 0.37 \quad 0.19 \quad 0.19 \quad 0.54 \quad 0.46 \quad 0.30 \quad 0.37)$. Here, $\mathbf{B}$ is rescaled to have columns of length 1. The unique variance of zero is a boundary solution. This may also occur for other models and estimation methods; see e.g. Bentler and Lee (1979). For the Candecomp/Parafac decomposition $(\mathbf{C}, \mathbf{B}, \mathbf{T})$ we have $k_{\mathbf{C}} = k_{\mathbf{B}} = k_{\mathbf{T}} = 2$. Hence, by condition (4.8) it is unique up to permutation and scaling. The eigenvalues of $\mathbf{\Sigma} - \mathbf{U}$ are all nonnegative: 0, 0, 0, 0.18, 0.50, 0.73, 1.02, 3.14. The percentage of explained common variance is 74.56. The percentages for each trait-method combination are given in Table 4.2. Note that the unique variances in $\mathbf{U}$ are much lower than in the solution of Bentler and Lee (1978). The model has 11 parameters and the sum-of-squares of $\mathbf{\Sigma} - \mathbf{U} - (\mathbf{C} \odot \mathbf{B})\,\mathbf{\Phi}\,(\mathbf{C} \odot \mathbf{B})^T$ equals 0.84. This is larger than for the solution of Bentler and Lee (1978), but we estimate under the restriction that $\mathbf{\Sigma} - \mathbf{U}$ is a covariance matrix.

For the interpretation of (4.15) we compute $(\mathbf{C} \odot \mathbf{B})\mathbf{\Phi}$; see Table 4.3. The first factor can be interpreted as leadership and extraversion (variables 3 and 4), and

Table 4.2: Percentages of estimated common variances (ECV%) and communalities for the Candecomp/Parafac covariance model (4.9) with $R = 2$ factors fitted to the MTMM data from Bentler and Lee (1978).

variable	ECV% oblique	ECV% orthogonal	ECV% Paralind	communalities
Am, self	98	97	89	1.00
At, self	36	35	36	0.63
Le, self	81	81	83	0.81
Ex, self	87	88	89	0.81
Am, peer	55	58	63	0.46
At, peer	37	36	35	0.54
Le, peer	88	86	82	0.70
Ex, peer	84	82	78	0.63

is similar to the second factor of the solution (4.14). The second factor can be interpreted as ambition, but only for self-report. This factor represents a difference in ambition measured by self-report and ambition measured by peer-report. This information cannot be obtained from (4.14) where both factors have the same weight for each condition (i.e., matrix $\mathbf{C}$ has only one column). The first factor in (4.14) is also interpreted as ambition, but there is no clear difference between methods. The larger weights in $\mathbf{C}$ for self-report in (4.14) also occur in $\mathbf{C}$ in (4.15), but are much more pronounced for the second factor. Note that in (4.15), the unique variance of ambition measured by self-report equals zero. The variance of this variable-method combination is almost completely explained in the covariance matrix $(\mathbf{C} \odot \mathbf{B})\mathbf{\Phi}(\mathbf{C} \odot \mathbf{B})^T$; see Table 4.2. The columns of $\mathbf{B}$ in (4.15) are somewhat similar to those of $\mathbf{B}$ in (4.14), except they are in reversed order, and the zero entry after rotation is not present in (4.15). The factors correlate -0.53 as opposed to being uncorrelated in (4.14).

Table 4.3: Estimated values of $(\mathbf{C} \odot \mathbf{B})\,\mathbf{\Phi}$ for the Candecomp/Parafac covariance model (4.9) with $R = 2$ factors fitted to the MTMM data from Bentler and Lee (1978).

variable	oblique		orthogonal		Paralind	
Am, self	-0.00	**0.84**	0.26	**0.97**	**0.85**	0
At, self	0.40	-0.02	0.45	0.15	0.24	0.45
Le, self	**0.75**	-0.09	**0.80**	0.16	0.40	**0.79**
Ex, self	**0.80**	-0.28	**0.80**	-0.01	0.26	**0.83**
Am, peer	0.26	0.23	0.25	0.40	**0.67**	0
At, peer	0.42	-0.12	0.43	0.06	0.18	0.35
Le, peer	**0.75**	-0.26	**0.76**	0.07	0.31	**0.62**
Ex, peer	**0.75**	-0.33	**0.76**	-0.00	0.20	**0.65**

Note: Numbers larger than 0.6 are in boldfont.

We also estimate the Candecomp/Parafac covariance model (4.9) with $R = 2$ orthogonal factors ($\mathbf{\Phi} = \mathbf{I}_2$). The results are as follows:

$$\mathbf{C} = \begin{pmatrix} 1.25 & 1.00 \\ 1.19 & 0.41 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 0.21 & 0.97 \\ 0.36 & 0.15 \\ 0.64 & 0.17 \\ 0.64 & -0.01 \end{pmatrix}, \tag{4.16}$$

where $\mathbf{B}$ has columns of length 1. Again, the Candecomp/Parafac decomposition $(\mathbf{C}, \mathbf{B}, \mathbf{T})$ is unique up to permutation and scaling by condition (4.8), since $k_{\mathbf{C}} = k_{\mathbf{B}} = k_{\mathbf{T}} = 2$. The percentage of explained common variance equals 74.16, of which 53.4 percent is due to factor one and 20.8 percent is due to factor two; see (4.13). The model has 10 parameters and the sum-of-squares of $\mathbf{\Sigma} - \mathbf{U} - (\mathbf{C} \odot \mathbf{B})(\mathbf{C} \odot \mathbf{B})^T$ equals 0.95. This is larger than for estimation with

two oblique factors, since we use the restriction of orthogonal factors here.

We use the values of $(\mathbf{C} \odot \mathbf{B})$ to interpret (4.16); see Table 4.3. As (4.15), the two factors are clearly related to leadership and extraversion (factor one) and ambition (factor two). Matrix $\mathbf{C}$ is similar to (4.15), but now the difference in weights for self-report and peer-report pertains almost only to factor two. With $\mathbf{\Phi} = \mathbf{I}_2$, one only has to use $\mathbf{C}$ and $\mathbf{B}$ to interpret the factors. Compared to the solution (4.14) of Bentler and Lee (1978), which also features two orthogonal factors, solution (4.16) yields a unique and proper solution, is easier to interpret, and contains more information.

Finally, we estimate a solution for the Candecomp/Parafac covariance model with two orthogonal factors and $\mathbf{C} = [\mathbf{c}\ \mathbf{c}]$. This is a Paralind model, as explained at the end of section 4.2. We have $(\mathbf{C} \odot \mathbf{B}) = (\mathbf{c} \otimes \mathbf{B})$, which implies that we actually have the same model as in Bentler and Lee (1978). Since the solution is not unique, we rotate $\mathbf{B}$ such that its (1,2) entry is zero, and rescale $\mathbf{C}$ such that its first entry equals 1. In this way, we obtain a solution that can be compared directly to (4.14). The result is:

$$\mathbf{c} = \begin{pmatrix} 1 \\ 0.78 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 0.85 & 0 \\ 0.24 & 0.45 \\ 0.40 & 0.79 \\ 0.26 & 0.83 \end{pmatrix},$$

which looks a lot like (4.14). However, the differences are that in our solution $\mathbf{\Sigma} - \mathbf{U}$ is a covariance matrix, and the unique variances in $\mathbf{U}$ are much smaller. The percentage of explained common variance equals 72.76. The model has 8 parameters and the sum-of-squares of $\mathbf{\Sigma} - \mathbf{U} - (\mathbf{c} \otimes \mathbf{B})(\mathbf{c} \otimes \mathbf{B})^T$ equals 1.09. This is larger than for the model with two orthogonal factors, since we use the restriction $\mathbf{C} = [\mathbf{c}\ \mathbf{c}]$.

4.3.2 Belief in a just world items of Lipkus et al. (1996)

To measure people's belief in a just world, Lipkus et al. (1996) argue that it is important to distinguish between seeing the world as just or unjust for oneself and seeing the world as just or unjust for others. They administered a belief in a just world scale containing eight items, where each item is formulated both as "for yourself" and as "for others". The items for yourself are listed below. The items for others are the same, except that the words "me" and "I" are replaced by "people" and "they" (and the sentence is grammatically corrected).

1. I feel that I get what I am entitled to have.
2. I feel that my efforts are noticed and rewarded.
3. I feel that people treat me fairly in life.
4. I feel that I earn the rewards and punishments I get.
5. I feel that when I meet with misfortune, I have brought it upon myself.
6. I feel that I get what I deserve.
7. I feel that people treat me with the respect I deserve.
8. I feel that the world treats me fairly.

Each item is answered on a 7-point scale, where 1 represents "totally disagree" and 7 represents "totally agree". As part of a larger study, the items of Lipkus et al. (1996) were administered in September 2012 in an online survey containing 246 paid American respondents aged 18 to 82. After deleting persons with missing data, $N = 236$ persons are kept. Note that we have $J = 8$ items and $K = 2$ conditions (yourself/others). The 16×16 correlation matrix $\mathbf{\Sigma}$ is given in the appendix.

Next, we estimate the Candecomp/Parafac covariance model (4.9), using the estimation procedure in section 4.2. We use orthogonal factors to ease the interpretation of the solution. The solution with $R = 2$ factors yields one general factor and one "for yourself" factor, and has 74.51 percent explained common variance. In the solution with $R = 3$, the factors have a more interesting interpretation. The results for $R = 3$ are:

$$\mathbf{C} = \begin{pmatrix} 1.78 & 1.32 & -0.68 \\ 1.35 & 1.62 & 1.13 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 0.30 & 0.35 & 0.16 \\ 0.44 & 0.24 & 0.29 \\ 0.45 & 0.17 & 0.48 \\ 0.28 & 0.44 & 0.25 \\ -0.02 & 0.56 & 0.18 \\ 0.28 & 0.46 & 0.30 \\ 0.42 & 0.16 & 0.49 \\ 0.42 & 0.23 & 0.50 \end{pmatrix},$$

$$\begin{aligned} \text{diag}(\mathbf{U}) = &(0.34 \;\; 0.13 \;\; 0.16 \;\; 0.19 \;\; 0.29 \;\; 0.15 \;\; 0.21 \;\; 0.11 \;\; 0.21 \;\; 0.25 \;\; 0.15 \\ &0.23 \;\; 0 \;\; 0.11 \;\; 0.19 \;\; 0.11)\,. \end{aligned} \tag{4.17}$$

Again, $\mathbf{B}$ is rescaled to have columns of length 1. The Candecomp/Parafac decomposition $(\mathbf{C}, \mathbf{B}, \mathbf{T})$ is unique up to permutation and scaling by condition (4.8), since $k_{\mathbf{C}} = 2$ and $k_{\mathbf{B}} = k_{\mathbf{T}} = 3$. The eigenvalues of $\mathbf{\Sigma} - \mathbf{U}$ are all nonnegative: 0, 0, 0, 0.01, 0.02, 0.08, 0.11, 0.17, 0.18, 0.27, 0.28, 0.33, 0.53, 1.24, 2.07, 7.88. The percentage of explained common variance is 84.54, where 38.0 percent is due to factor 1, 33.3 percent is due to factor 2, and 13.2 percent is due to factor 3. Note that the fit of the model is quite good, with high explained common variance and low unique variances in $\mathbf{U}$. The sum-of-squares of $\mathbf{\Sigma} - \mathbf{U} - (\mathbf{C} \odot \mathbf{B})(\mathbf{C} \odot \mathbf{B})^T$ equals only 1.06.

For the interpretation of the solution (4.17) we compute $(\mathbf{C} \odot \mathbf{B})$, which can be found in Table 4.4. Factor 1 is a mixture of especially items 2, 3, 7, 8, and has somewhat smaller loadings for especially items 1, 4, 6. The loadings are higher for the self condition than for others. Factor 2 is a mixture of items 1, 4, 5, 6, and has higher loadings for others than for the self condition. One could say that the items of factor 1 pertain to a more implicit or passive idea of justice (e.g., how you are treated), while the items of factor 2 pertain to a more explicit or active idea of justice (e.g., what you get). The third factor represents a contrast between the two conditions. Its loadings are largest for items 3, 7, and 8 for the others condition. The presence of this contrast factor is in line with the substantive arguments in Lipkus et al. (1996) for distinguishing the two conditions (yourself/others) in belief in a just world scales.

Table 4.4: Estimated values of $(\mathbf{C} \odot \mathbf{B})$ for the Candecomp/Parafac covariance model (4.9) with $R = 3$ orthogonal factors fitted to the belief in a just world data.

item	factor 1	factor 2	factor 3	item	factor 1	factor 2	factor 3
1, self	**0.53**	0.47	-0.11	1, others	0.40	**0.57**	0.18
2, self	**0.79**	0.31	-0.19	2, others	**0.60**	0.38	0.32
3, self	**0.80**	0.23	-0.32	3, others	**0.61**	0.28	**0.54**
4, self	**0.50**	**0.58**	-0.17	4, others	0.38	**0.71**	0.28
5, self	-0.04	**0.73**	-0.12	5, others	-0.03	**0.90**	0.21
6, self	0.49	**0.60**	-0.20	6, others	0.37	**0.74**	0.34
7, self	**0.76**	0.21	-0.33	7, others	**0.57**	0.25	**0.55**
8, self	**0.75**	0.31	-0.33	8, others	**0.57**	0.37	**0.56**

Note: Numbers larger than 0.5 are in boldfont.

4.4 Simulation study

Here, we assess the performance of the estimation procedure outlined in section 4.2 to retrieve underlying factors in simulated three-mode data. For given $\mathbf{C}, \mathbf{B}, \mathbf{\Phi}, \mathbf{U}$, we create random data $\mathbf{X}_{(N\times JK)}$ with population correlation matrix $\mathbf{\Sigma}$ in (4.9). Next, we apply the estimation procedure of section 4.2 on the sample covariance matrix of $\mathbf{X}_{(N\times JK)}$, and check whether the underlying factors are retrieved. Below, we explain the details of our simulations.

We consider the cases of $R = 2$ and $R = 3$ factors. The true matrices $\mathbf{C}$, $\mathbf{B}$, and $\mathbf{\Phi}$ are the following. For $R = 2$, we use

$$\mathbf{C}_1 = \begin{pmatrix} 1.00 & 0.80 \\ 0.80 & 1.20 \end{pmatrix}, \qquad \mathbf{\Phi}_1 = \begin{pmatrix} 1 & -0.40 \\ -0.40 & 1 \end{pmatrix},$$

and

$$\mathbf{B}_1 = \begin{pmatrix} 0.80 & 0.10 \\ 0.10 & 0.83 \\ 0.83 & 0.10 \\ 0.10 & 0.79 \\ 0.83 & 0.10 \\ 0.10 & 0.82 \end{pmatrix}, \qquad \mathbf{B}_2 = \begin{pmatrix} 0.65 & 0.10 \\ 0.10 & 0.63 \\ 0.61 & 0.10 \\ 0.10 & 0.70 \\ 0.64 & 0.10 \\ 0.10 & 0.62 \end{pmatrix}.$$

Hence, we use one true matrix $\mathbf{C}_1$, and two true matrices $\mathbf{B}_1$ and $\mathbf{B}_2$. In the latter matrices, each item loads high on one factor only, and each factor has high loadings from three items. Matrix $\mathbf{B}_1$ has higher loadings than matrix $\mathbf{B}_2$. We consider both orthogonal factors ($\mathbf{\Phi} = \mathbf{I}_2$) and oblique factors ($\mathbf{\Phi} = \mathbf{\Phi}_1$). For $R = 3$, we take the true $\mathbf{B}$ and $\mathbf{C}$ from the solution (4.17) for the belief in a just world scales. The three factors are orthogonal.

After the true $\mathbf{C}$ and $\mathbf{B}$ are chosen, the unique variances on the diagonal of $\mathbf{U}$ are determined such that the population correlation matrix $\mathbf{\Sigma}$ in (4.9) has ones on the diagonal. For the sample size N, we consider $N = 100$ and $N = 500$. The data are generated as

$$\mathbf{X}_{(N\times JK)} = \mathbf{Z}_{(N\times JK)}\,\mathbf{\Sigma}^{1/2}\,, \tag{4.18}$$

where $\mathbf{Z}_{(N\times JK)}$ has random entries from the standard normal distribution, and $\mathbf{\Sigma}$ is the population correlation matrix. For each choice of true model and sample size, we generate 100 datasets (4.18), and fit the Candecomp/Parafac covariance model (4.9) to the sample covariance matrix.

To assess factor retrieval, we compare the true values of the loadings in matrices $\mathbf{C}$ and $\mathbf{B}$ to their estimates. For this, we use the congruence coefficient of the true value and estimate of each column of $\mathbf{C}$ and $\mathbf{B}$. For two vectors $\mathbf{h}_1$ and $\mathbf{h}_2$, the congruence coefficient is given by

$$\frac{\mathbf{h}_1^T\mathbf{h}_2}{\sqrt{\mathbf{h}_1^T\mathbf{h}_1}\,\sqrt{\mathbf{h}_2^T\mathbf{h}_2}}\,.$$

For two-mode factor analysis, congruence coefficients of the columns of two loading matrices are used as a measure of factor similarity. Absolute values of 0.85 to 0.94 correspond to fair similarity, while an absolute value higher than 0.95 implies equal factors (Lorenzo-Seva & Ten Berge, 2006). Note that by using congruence coefficients, we do not need to deal with the scaling indeterminacy of the columns of $\mathbf{C}$ and $\mathbf{B}$.

In Tables 4.5 and 4.6 below, we report the mean and standard deviation of the congruence coefficients of the columns of $\mathbf{C}$ and $\mathbf{B}$ for each case. Also the average communality is given for each case. In each case we estimate the model using both orthogonal factors and oblique factors. We only report the results

for sample size $N = 100$. For $N = 500$ the congruence coefficients are slightly larger and have slightly smaller standard deviation.

To get an idea of the estimation accuracy of the values of the loadings in $\mathbf{B}$ and $\mathbf{C}$, we rescale the columns of the estimated $\mathbf{B}$ and $\mathbf{C}$ to match the sum-of-squares of the columns of the true $\mathbf{B}$ and $\mathbf{C}$. For each rescaled estimated $\mathbf{B}$ and $\mathbf{C}$, we compute the mean absolute deviation and the mean bias, which are defined for $\mathbf{B}$ as

$$\text{MAD}(\hat{\mathbf{B}}, \mathbf{B}^{(\text{true})}) = \frac{\sum_{j=1}^{J}\sum_{r=1}^{R} |\hat{b}_{jr} - b_{jr}^{(\text{true})}|}{JR},$$

$$\text{BIAS}(\hat{\mathbf{B}}, \mathbf{B}^{(\text{true})}) = \frac{\sum_{j=1}^{J}\sum_{r=1}^{R} (\hat{b}_{jr} - b_{jr}^{(\text{true})})}{JR},$$

and analogously for $\mathbf{C}$. We also report the estimation accuracy of the unique variances $\mathbf{U}$ in terms of MAD and BIAS. In Tables 4.7 and 4.8, we report the MAD and BIAS values for $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{U}$.

For $R = 2$ the recovery of the true loadings is very good in general. For larger communalities, the recovery is clearly better. Also, using orthogonal factors in the estimation improves the recovery when the true model has orthogonal factors. For the MAD values this is only true for $\mathbf{B}$. The BIAS of $\mathbf{C}$ is close to zero, while the estimates of $\mathbf{B}$ are slightly negatively biased. Note that using orthogonal factors in the estimation while the true model has oblique factors still results in relatively high congruence coefficients. The estimation of the unique variances in $\mathbf{U}$ is less good. The MAD values are rather large and the estimates are negatively biased.

Table 4.5: Congruence coefficients (mean and standard deviation, rounded to two decimals) for the simulated cases with $R = 2$ factors and sample size $N = 100$.

$\mathbf{C}$	$\mathbf{B}$	$\mathbf{\Phi}$	comm.	estimation	congr.coeff. $\mathbf{C}$		congr.coeff. $\mathbf{B}$		div.comp.
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	0.63	orthogonal	1.00 (0.00)	1.00 (0.00)	0.99 (0.01)	0.99 (0.01)	-
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	0.63	oblique	1.00 (0.00)	1.00 (0.00)	0.99 (0.01)	0.99 (0.01)	0
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	0.39	orthogonal	1.00 (0.00)	1.00 (0.00)	0.97 (0.04)	0.98 (0.03)	-
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	0.39	oblique	1.00 (0.00)	1.00 (0.00)	0.96 (0.07)	0.97 (0.04)	0
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	0.57	orthogonal	1.00 (0.00)	1.00 (0.00)	0.97 (0.03)	0.97 (0.02)	-
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	0.57	oblique	1.00 (0.00)	1.00 (0.00)	0.98 (0.03)	0.98 (0.02)	2
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	0.35	orthogonal	1.00 (0.00)	1.00 (0.00)	0.95 (0.05)	0.96 (0.04)	-
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	0.35	oblique	1.00 (0.00)	1.00 (0.00)	0.94 (0.07)	0.96 (0.05)	2

Table 4.6: Congruence coeffcients (mean and standard deviation, rounded to two decimals) for true $\mathbf{C}$ and $\mathbf{B}$ in (4.17), $R = 3$ orthogonal factors and sample size $N = 100$. Average communality is 0.69.

estimation	congr.coeff. $\mathbf{C}$			congr.coeff. $\mathbf{B}$			div.comp.
orthogonal	1.00 (0.00)	1.00 (0.00)	0.98 (0.03)	0.99 (0.02)	0.98 (0.03)	0.99 (0.01)	-
oblique	1.00 (0.01)	1.00 (0.00)	0.98 (0.04)	0.98 (0.02)	0.94 (0.07)	0.99 (0.01)	1

Table 4.7: Mean and standard deviation of MAD and BIAS between estimated and true values of **B**, **C**, and **U** for the case of $R = 2$ and sample size $N = 100$

C	**B**	**Φ**	estimation	MAD			BIAS		
				C	**B**	**U**	**C**	**B**	**U**
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	orthogonal	0.03	0.05	0.13	0.00	0.00	-0.13
				(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.02)
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	oblique	0.02	0.06	0.13	0.00	0.00	-0.12
				(0.01)	(0.02)	(0.02)	(0.00)	(0.03)	(0.02)
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	orthogonal	0.06	0.07	0.23	0.00	-0.01	-0.22
				(0.03)	(0.03)	(0.03)	(0.01)	(0.02)	(0.03)
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	oblique	0.04	0.08	0.22	0.00	-0.01	-0.22
				(0.02)	(0.03)	(0.03)	(0.01)	(0.04)	(0.03)
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	orthogonal	0.03	0.10	0.15	0.00	-0.08	-0.15
				(0.01)	(0.02)	(0.02)	(0.00)	(0.02)	(0.02)
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	oblique	0.03	0.07	0.15	0.00	0.00	-0.15
				(0.02)	(0.02)	(0.02)	(0.00)	(0.04)	(0.02)
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	orthogonal	0.05	0.10	0.24	0.00	-0.06	-0.23
				(0.03)	(0.03)	(0.02)	(0.00)	(0.02)	(0.03)
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	oblique	0.06	0.10	0.24	0.00	-0.01	-0.23
				(0.03)	(0.04)	(0.03)	(0.00)	(0.04)	(0.03)

Table 4.8: Mean and standard deviation of MAD and BIAS between estimated and true values of **B**, **C**, and **U** for the case of $R = 3$ and sample size $N = 100$.

estimation	MAD			BIAS		
	C	**B**	**U**	**C**	**B**	**U**
orthogonal	0.09	0.03	0.13	0.00	0.00	-0.13
	(0.07)	(0.01)	(0.01)	(0.07)	(0.00)	(0.01)
oblique	0.20	0.05	0.13	-0.07	0.00	-0.13
	(0.21)	(0.02)	(0.01)	(0.19)	(0.01)	(0.01)

For $R = 3$ the recovery results are also very good. For oblique estimation, however, the MAD and BIAS values for $\mathbf{C}$ are somewhat larger. After inspection of the estimation results, it turns out that in 22 of the 100 replications two of the three factors collapse into one factor and this makes the estimated CP decomposition nonunique. This nonuniqueness mainly affects the estimated coefficients in $\mathbf{C}$. Perhaps this estimation behavior is encouraged by the large congruence coefficient (0.97) of the first two columns of the true $\mathbf{C}$. Naturally, this does not happen for orthogonal estimation. As for $R = 2$, the estimation of $\mathbf{U}$ is negatively biased.

When oblique factors are used in the estimation, we also report the number of cases diverging components are encountered in the Candecomp/Parafac algorithm. Two components are labeled diverging when their congruence coefficient is smaller than -0.90; see e.g. Stegeman (2012). The occurrence of diverging components often indicates nonexistence of a best-fitting Candecomp/Parafac model. As can be seen, there were only few such cases. For sample size $N = 500$ no cases of diverging components were encountered.

4.5 Discussion

In this chapter, we have proposed and demonstrated a method for three-mode factor analysis using MRFA to estimate unique variances $\mathbf{U}$ and Candecomp/Parafac to estimate the covariance matrix of the common part. By using MRFA, the matrix $\mathbf{\Sigma} - \mathbf{U}$ is guaranteed to be a covariance matrix. This makes it possible to compute the percentage of explained common variance. For other methods of (two-mode or three-mode) factor analysis, this is often not possible. Also, our factor correlation matrix $\mathbf{\Phi}$ is guaranteed to be a covariance matrix.

By using the Candecomp/Parafac covariance model, the factors and weight

matrices are unique up to permutation and scaling under mild conditions. This is not the case for covariance models based upon the three-mode model by Tucker (1966). Also, identifiability is often hard to prove for confirmatory factor analysis models. Besides yielding a unique solution, the Candecomp/Parafac covariance model is parsimonious and easy to interpret.

The simulation study shows that our relatively simple estimation procedure performs very well in retrieving underlying factors when the data is randomly sampled from a normal distribution with a covariance matrix $\mathbf{\Sigma}$ satisfying the Candecomp/Parafac covariance model. However, the simulations for $R = 3$ with oblique estimation show that diverging components may hamper estimation accuracy. For orthogonal factors, this is not possible.

To select the number of factors in our Candecomp/Parafac covariance model, one may use the criterion of increase in explained common variance. When adding an additional factor increases the percentage of explained common variance only little, one may decide not to include an additional factor in the analysis. However, one should also consider the added value of the additional factor in terms of interpretation.

Although we have focused on the Candecomp/Parafac covariance model, our approach can be generalized to any model for the common covariance part. That is, after estimating $\mathbf{U}$ by MRFA, any model can be fit to $\mathbf{\Sigma} - \mathbf{U}$. As mentioned in sections 4.1.2 and 4.2, it may be convenient to apply a Tucker-based covariance model instead of Candecomp/Parafac. To determine whether a Candecomp/Parafac or Tucker-type model is appropriate, and which number(s) of factors should be used, Ceulemans and Kiers (2006) have proposed a method based on selecting the model with the highest fit for an acceptable number of free parameters. Naturally, also the size of the interaction terms in a Tucker-type model, the desirability of rotational nonuniqueness, and ease of interpretation

should be taken into account when selecting an appropriate model.

Appendix: correlation matrix of Belief in a just world data

	for yourself								for others							
item	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
1	1	0.56	0.46	0.51	0.25	0.53	0.42	0.52	0.61	0.50	0.40	0.46	0.29	0.52	0.39	0.41
2	0.56	1	0.75	0.63	0.24	0.55	0.68	0.72	0.40	0.50	0.49	0.37	0.12	0.38	0.35	0.38
3	0.46	0.75	1	0.60	0.19	0.51	0.77	0.80	0.33	0.44	0.37	0.32	0.05	0.31	0.29	0.27
4	0.51	0.63	0.60	1	0.47	0.67	0.52	0.61	0.50	0.51	0.41	0.61	0.33	0.54	0.34	0.37
5	0.25	0.24	0.19	0.47	1	0.56	0.16	0.30	0.35	0.30	0.28	0.47	0.65	0.41	0.22	0.27
6	0.53	0.55	0.51	0.67	0.56	1	0.51	0.65	0.47	0.48	0.43	0.55	0.40	0.61	0.35	0.39
7	0.42	0.68	0.77	0.52	0.16	0.51	1	0.73	0.35	0.43	0.35	0.31	0.04	0.34	0.34	0.31
8	0.52	0.72	0.80	0.61	0.30	0.65	0.73	1	0.37	0.49	0.43	0.40	0.18	0.46	0.36	0.42
1	0.61	0.40	0.33	0.50	0.35	0.47	0.35	0.37	1	0.62	0.46	0.62	0.52	0.66	0.50	0.54
2	0.50	0.50	0.44	0.51	0.30	0.48	0.43	0.49	0.62	1	0.67	0.61	0.34	0.65	0.57	0.57
3	0.40	0.49	0.37	0.41	0.28	0.43	0.35	0.43	0.46	0.67	1	0.55	0.35	0.59	0.74	0.73
4	0.46	0.37	0.32	0.61	0.47	0.55	0.31	0.40	0.62	0.61	0.55	1	0.62	0.77	0.51	0.56
5	0.29	0.12	0.05	0.33	0.65	0.40	0.04	0.18	0.52	0.34	0.35	0.62	1	0.62	0.36	0.40
6	0.52	0.38	0.31	0.54	0.41	0.61	0.34	0.46	0.66	0.65	0.59	0.77	0.62	1	0.57	0.65
7	0.39	0.35	0.29	0.34	0.22	0.35	0.34	0.36	0.50	0.57	0.74	0.51	0.36	0.57	1	0.77
8	0.41	0.38	0.27	0.37	0.27	0.39	0.31	0.42	0.54	0.57	0.73	0.56	0.40	0.65	0.77	1

Note: Data provided by K. Stroebe, Department of Social Psychology, University of Groningen.

Chapter 5

Three-mode factor analysis by means of Tucker3

5.1 Introduction

In this chapter we discuss a different approach to three-mode factor analysis that is based on the Tucker3 model of Tucker (1966). Unlike CP, in the Tucker3 model, each mode of the data may have a different number of components, and all their interaction strengths are included as entries g_{rpq} of the so called *core array*. Suppose that we have R components for the N observations, P components for the J variables, and Q components for the K conditions. The Tucker3 model can then be written in the form of

$$\mathbf{X}_{(N\times JK)} = \mathbf{F}\,\mathbf{G}\,(\mathbf{C}\otimes\mathbf{B})^T + \mathbf{E}_{(N\times JK)}\,, \tag{5.1}$$

with matrix $\mathbf{G}$ $(R \times PQ)$ of interaction strengths as follows.

$$\mathbf{G} = \left[\begin{array}{ccc|c|ccc} g_{111} & \cdots & g_{1P1} & \cdots & g_{11Q} & \cdots & g_{1PQ} \\ \vdots & & \vdots & & \vdots & & \vdots \\ g_{R11} & \cdots & g_{RP1} & \cdots & g_{R1Q} & \cdots & g_{RPQ} \end{array}\right].$$

It is well known that the three-mode Tucker model is not unique. All of $\mathbf{F}$, $\mathbf{C}$, and $\mathbf{B}$ can be rotated, with inverse transformations applied to the interaction strengths in $\mathbf{G}$, without affecting the model part $\mathbf{FG}(\mathbf{C} \otimes \mathbf{B})^T$; see section 2.2.6. The covariance model corresponding to (5.1) is given as

$$\mathbf{\Sigma} = (\mathbf{C} \otimes \mathbf{B})\, \mathbf{G}^T\, \mathbf{\Phi}\, \mathbf{G}\, (\mathbf{C} \otimes \mathbf{B})^T + \mathbf{U}, \tag{5.2}$$

where $\mathbf{C}$ is an $K \times Q$ matrix containing the Q method components as columns, $\mathbf{B}$ is a $J \times P$ matrix containing the P variable components as columns, $\mathbf{C} \otimes \mathbf{B}$ is the Kronecker product of $\mathbf{C}$ and $\mathbf{B}$.

For simplicity, one may consider $\mathbf{\Psi} = \mathbf{G}^T\mathbf{\Phi}\mathbf{G}$ as factor covariance matrix. The case of diagonal $\mathbf{\Psi}$ can be rewritten as $\mathbf{\Sigma} = (\mathbf{C}\mathbf{C}^T \otimes \mathbf{B}\mathbf{B}^T) + \mathbf{U}$ and is known as the direct product model; see Browne (1984) or Wothke (1996). In this model, $\mathbf{C}\mathbf{C}^T$ and $\mathbf{B}\mathbf{B}^T$ may be seen as covariance matrices corresponding to methods and variables, respectively.

Fitting the covariance model (5.2) can be done by directly fitting the component model (5.1) to the data. Alternating least squares algorithms minimizing the sum-of-squares of $\mathbf{E}_{(N \times JK)}$ can be found in Kroonenberg and De Leeuw (1980) and Kiers et al. (1992), see section 2.2.5. Contrary to the confirmatory factor analysis approach, convergence problems do not often occur. One may also fit the covariance model (5.2) by an algorithm for nonlinear optimization. For example, Bentler and Lee (1978, 1979) propose to use a Gauss-Newton algorithm. For an overview of three-mode component and factor models based

on Tucker (1966), we refer to Kroonenberg and Oort (2003). For an accessible introduction to three-mode component analysis, see Kiers and Van Mechelen (2001).

In section 5.2, we provide an estimation procedure to fit model (5.2) such that $\mathbf{\Sigma} - \mathbf{U}$ is a covariance matrix. This allows for the computation of the percentage of explained common variance. The estimation procedure is analogous to section 4.2 for three-mode factor analysis by means of CP. In sections 5.3 and 5.4, we apply model (5.2) to datasets in the literature. Finally, section 5.5 contains a discussion of our findings.

5.2 Estimation Procedure

Here, we present our estimation procedure for the Tucker3 covariance model (5.2). After having computed the data covariance matrix $\mathbf{\Sigma}$, the steps of our estimation procedure for model (5.2) are as follows.

1. Use the MRFA algorithm of Ten Berge and Kiers (1991) to estimate $\mathbf{U}$. This implies that $\mathbf{U}$ is nonnegative, $\mathbf{\Sigma} - \mathbf{U}$ is a covariance matrix, and the trace of $\mathbf{\Sigma} - \mathbf{L} - \mathbf{U}$ is minimal, where $\mathbf{L}$ is a best rank-PQ approximation of $\mathbf{\Sigma} - \mathbf{U}$.

2. Compute the eigendecomposition $\mathbf{\Sigma} - \mathbf{U} = \mathbf{V}\,\mathbf{S}\,\mathbf{V}^T$, with $\mathbf{V}$ having orthonormal columns, and $\mathbf{S}$ the diagonal matrix containing the eigenvalues in decreasing order. This is also the singular value decomposition of $\mathbf{\Sigma} - \mathbf{U}$. Let $\mathbf{P} = \mathbf{V}\mathbf{S}^{1/2}$, which implies $\mathbf{\Sigma} - \mathbf{U} = \mathbf{P}\mathbf{P}^T$.

3. Fit the Tucker3 model as $\mathbf{P} \approx (\mathbf{C} \otimes \mathbf{B})\mathbf{G}^T\mathbf{T}^T$ by using the alternating least squares algorithm of Kroonenberg and De Leeuw (1980) with number of components in each mode respectively R, P, Q. Matrix $\mathbf{P}$ ($JK \times JK$) is

the matricized $K \times J \times JK$ array, $\mathbf{G}$ $(R \times PQ)$ is the matricized $R \times P \times Q$ core array, and $\mathbf{T}$ is a $JK \times R$ matrix satisfying $\mathbf{T}^T\mathbf{T} = \mathbf{I}_R$. The columns of $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{T}$ are scaled such that they have length 1. We obtain $\mathbf{\Sigma} - \mathbf{U} \approx (\mathbf{C} \otimes \mathbf{B})\mathbf{\Psi}(\mathbf{C} \otimes \mathbf{B})^T$, where $\mathbf{\Psi} = \mathbf{G}^T\mathbf{T}^T\mathbf{T}\mathbf{G} = \mathbf{G}^T\mathbf{G}$ is an *interaction matrix* of the factors in the second and third modes. We set $R = PQ$ to not constrain the rank of $\mathbf{\Psi}$. We use Joint Orthomax rotation of (Kiers, 1998a) to rotate the Tucker3 solution to simple structure in $\mathbf{B}$, $\mathbf{C}$, and in the core $\mathbf{G}$. We use the standard weights specified in Kiers (1998a); see also section 3.6.2. We evaluate the Tucker3 fit as

$$100 - 100 \cdot \frac{\text{ssq}(\mathbf{P} - (\mathbf{C} \otimes \mathbf{B})\mathbf{G}^T\mathbf{T}^T)}{\text{ssq}(\mathbf{P})}, \tag{5.3}$$

which is the percentage of the sum-of-squares of $\mathbf{P}$ that is fitted by $(\mathbf{C} \otimes \mathbf{B})\mathbf{G}^T\mathbf{T}^T$. In the alternating least squares algorithm, $(\mathbf{C} \otimes \mathbf{B})\mathbf{G}^T\mathbf{T}^T$ is the regression of $\mathbf{P}$ on $(\mathbf{C} \otimes \mathbf{B})\mathbf{G}^T$. Since the regression and the residual are orthogonal, it follows that (5.3) is equal to

$$100 \cdot \text{ssq}((\mathbf{C} \otimes \mathbf{B})\mathbf{G}^T\mathbf{T}^T)/\text{ssq}(\mathbf{P}).$$

Since we were not able to construct an algorithm for the simultaneous estimation of $\mathbf{U}$ and $\mathbf{C}$, $\mathbf{B}$, $\mathbf{\Psi}$ under the restriction that $\mathbf{\Sigma} - \mathbf{U}$ is a covariance matrix, we instead estimate $\mathbf{U}$ and $\mathbf{C}$, $\mathbf{B}$, $\mathbf{\Psi}$ sequentially. First, we estimate $\mathbf{U}$ by MRFA based on a rank-PQ factor model for $\mathbf{\Sigma} - \mathbf{U}$. For fitting $(\mathbf{C} \otimes \mathbf{B})\mathbf{\Psi}(\mathbf{C} \otimes \mathbf{B})^T$ to $\mathbf{\Sigma} - \mathbf{U}$, we have chosen to fit Tucker3 to $\mathbf{P}$ with $\mathbf{\Sigma} - \mathbf{U} = \mathbf{P}\mathbf{P}^T$. We run the algorithm of Tucker3 that is programmed by Henk A. L. Kiers 10 times for random starting with convergence criterion $1e - 9$ and keep the solution with the highest fit percentage. The percentage of explained common variance in the Tucker3 covariance model is given by (5.3), which can be written as

$$100 \cdot \frac{\text{trace}((\mathbf{C} \otimes \mathbf{B})\mathbf{\Psi}(\mathbf{C} \otimes \mathbf{B})^T)}{\text{trace}(\mathbf{\Sigma} - \mathbf{U})}. \tag{5.4}$$

To obtain a percentage of explained common variance for each variable-condition combination separately, we proceed as follows. Contrary to the two-mode case of MRFA, it may happen that some diagonal entry of $(\mathbf{C} \otimes \mathbf{B})\mathbf{\Psi}(\mathbf{C} \otimes \mathbf{B})^T$ is larger than the corresponding communality on the diagonal of $\mathbf{\Sigma} - \mathbf{U}$. Because of this, we formulate the explained common variance per variable-condition combination analogous to (5.3) rather than to (5.4). A particular variable-condition corresponds to a row of $\mathbf{P} - (\mathbf{C} \otimes \mathbf{B})\mathbf{G}^T\mathbf{T}^T$. Let row m of this matrix be denoted as $\mathbf{q}_m^T$. Then we define the corresponding percentage of explained common variance as

$$100 - 100 \cdot \frac{\text{ssq}(\mathbf{q}_m^T)}{(\mathbf{\Sigma} - \mathbf{U})_{mm}},$$

where $(\mathbf{\Sigma} - \mathbf{U})_{mm}$ is the corresponding communality.

As mentioned in section 4.1.1, $(\mathbf{C} \otimes \mathbf{B})\mathbf{\Psi}$ is also the covariance matrix between the variables and factors. This matrix is used to interpret the factors.

5.3 Application I

In this section, we apply our model to a data set from Dickinson and Tice (1973). The correlation matrix of the Dickinson and Tice (1973) job behavior ratings is reproduced in Table 5.1. This data set is obtained from $N = 149$ subjects with $J = 3$ traits assessed by $K = 3$ methods. The three traits are: Getting along with others (G), Dedication (D), and Ability to Apply Learning (L). And the three methods are: Peer Nominations (PN), Peer Checklist ratings (PC), and Supervisor Checklist ratings (SC).

In section 5.3.1 we fit our three-mode CP factor model (4.9) to this data set. In section 5.3.2 we fit our three-mode Tucker3 factor model (5.2) to this data set and compare the results to those in section 5.3.1.

5.3.1 Three-mode CP factor analysis solution

Here, we fit our three-mode CP factor model (4.9) to this data set. Following is the three-mode CP factor solution with $R = 2$ orthogonal factors

$$\mathbf{B} = \begin{pmatrix} 0.64 & 0.74 \\ 0.62 & 0.58 \\ 0.45 & 0.34 \end{pmatrix} \begin{matrix} \text{G} \\ \text{D}, \\ \text{L} \end{matrix} \quad \mathbf{C} = \begin{pmatrix} 0.99 & 0.68 \\ 0.59 & -0.64 \\ 0.74 & -0.61 \end{pmatrix} \begin{matrix} \text{PN} \\ \text{PC}, \\ \text{SC} \end{matrix} \tag{5.5}$$

and the unique variance $\mathbf{U} = \text{diag}(0.18\ 0.38\ 0.59\ 0.49\ 0.47\ 0.67\ 0.15\ 0.55\ 0.68)$. Matrix $\mathbf{B}$ is rescaled to have columns of length 1. The percentage of explained common variance is 64.68%, where 38.81% is due to factor 1, and 25.87% is due to factor 2. The percentages of explained common variance for each trait-

Table 5.1: Correlations of three traits Getting along with others (G), Dedication (D), and Ability to Apply Learning (L) measured by three methods Peer Nominations (PN), Peer Checklist ratings (PC), and Supervisor Checklist ratings (SC).

	PN			PC			SC		
	G	D	L	G	D	L	G	D	L
G	1	0.524	0.241	0.071	0.022	0.076	0.136	-0.028	-0.054
D	0.524	1	0.403	0.102	0.096	0.102	0.132	0.168	0.162
L	0.241	0.403	1	-0.018	0.018	0.100	0.061	0.135	0.252
G	0.071	0.102	-0.018	1	0.435	0.342	0.243	0.093	0.053
D	0.022	0.096	0.018	0.435	1	0.347	0.203	0.209	0.108
L	0.076	0.102	0.100	0.342	0.347	1	0.100	0.042	0.108
G	0.136	0.132	0.061	0.243	0.203	0.100	1	0.461	0.294
D	-0.028	0.168	0.135	0.093	0.209	0.042	0.461	1	0.280
L	-0.054	0.162	0.252	0.053	0.108	0.108	0.294	0.280	1

Note: Data taken from Dickinson and Tice (1973)

method combination are given in column 8 of Table 5.4. They are between 44 and 91 percent, where the percentages of L-PC and L-SC are rather low (44% and 45%). The model has 10 parameters, not counting unique variances.

For interpretation of (5.5) we compute $(\mathbf{C} \odot \mathbf{B})\mathbf{\Phi}$; see columns 4 and 5 of Table 5.2. Factor 1 is a general factor with larger loadings for Peer Nominations (PN) and Supervisor Checklist ratings (SC). Factor 2 represents a contrast between Peer Nominations (PN) and Peer Checklist ratings (PC) and Supervisor Checklist ratings (SC), mostly for Getting along with others (G) and Dedication (D).

The solution with $R = 3$ orthogonal factors is:

$$\mathbf{B} = \begin{pmatrix} 0.63 & 0.79 & 0.61 \\ 0.63 & 0.52 & 0.61 \\ 0.46 & 0.30 & 0.49 \end{pmatrix} \begin{matrix} \text{G} \\ \text{D}, \\ \text{L} \end{matrix} \quad \mathbf{C} = \begin{pmatrix} 1.04 & -0.44 & -0.41 \\ 0.62 & 0.20 & 0.90 \\ 0.61 & 0.97 & -0.17 \end{pmatrix} \begin{matrix} \text{PN} \\ \text{PC}, \\ \text{SC} \end{matrix} \tag{5.6}$$

and the unique variance $\mathbf{U} = \text{diag}(0.24\ 0.36\ 0.60\ 0.49\ 0.47\ 0.67\ 0.08\ 0.56\ 0.67)$. Matrix $\mathbf{B}$ is rescaled to have columns of length 1. The percentage of explained common variance is 82.76%, where 37.60% is due to factor 1, 24.47% is due to factor 2, and 20.69% is due to factor 3. The percentages of explained common variance for each trait-method combination are given in column 10 of Table 5.4. They are between 59 and 94 percent, where the percentages of L-PC and L-SC are much larger than those in case $R = 2$. The model has 15 parameters.

For interpretation of (5.6) we compute $(\mathbf{C} \odot \mathbf{B})\mathbf{\Phi}$; see columns 9, 10, 11 of Table 5.2. Each factor loads high on one method. Factor 1 is a general factor with highest loadings for Peer Nominations (PN). Factor 2 can be interpreted as a combination between Getting along with others (G) and Dedication (D) for Supervisor Checklist ratings (SC). Factor 3 is also a general factor for Peer Checklist ratings (PC).

We also obtain similar matrices $\mathbf{B}$ in the case of oblique factors with $R = 2$ and $R = 3$. The oblique solutions yield highly correlated factors (0.69 for $R = 2$ and 0.84 for $R = 3$). This implies that the solutions with oblique factors give less clear interpretation than with orthogonal factors; see Table 5.2.

Table 5.2: Values of $(\mathbf{C} \odot \mathbf{B})\,\mathbf{\Phi}$ for the estimated CP covariance model (4.9) with $R = 2$ and $R = 3$ factors fitted to the MTMM data from Dickinson and Tice (1973).

variable	$R = 2$				$R = 3$					
	oblique		orthogonal		oblique			orthogonal		
G-PN	0.33	-0.32	**0.64**	**0.50**	0.38	0.14	-0.14	**0.65**	-0.35	-0.25
D-PN	**0.44**	-0.12	**0.61**	**0.40**	**0.46**	0.16	0.07	**0.65**	-0.23	-0.25
L-PN	0.32	-0.05	**0.44**	0.23	0.28	0.04	0.07	**0.47**	-0.13	-0.20
G-PC	**0.54**	**0.55**	0.37	**-0.46**	**0.59**	**0.68**	0.28	0.38	0.16	**0.55**
D-PC	**0.48**	**0.46**	0.36	-0.37	**0.58**	**0.68**	0.25	0.38	0.10	**0.55**
L-PC	0.32	0.31	0.26	-0.21	**0.46**	**0.54**	0.19	0.28	0.06	**0.44**
G-SC	**0.63**	**0.58**	**0.47**	**-0.45**	**0.54**	0.23	**0.85**	0.37	**0.77**	-0.10
D-SC	**0.57**	**0.50**	**0.45**	-0.35	0.40	0.15	**0.62**	0.38	**0.51**	-0.10
L-SC	0.38	0.33	0.33	-0.21	0.24	0.07	0.39	0.27	0.29	-0.08

Note: Numbers larger than 0.4 are in boldfont.

5.3.2 Three-mode Tucker3 factor analysis solution

Next, we fit our model (5.2) to the data set of Dickinson and Tice (1973). The explained common variances for $P \in \{1, 2\}$, and $Q \in \{2, 3\}$ are given in the following table.

P	Q	ECV% for Tucker3	ECV% for CP with $R = PQ$
1	2	64.41	64.68
2	2	76.11	
1	3	81.69	82.76
2	3	91.56	

Below, we present the solutions with $P = 1, P = 2$, $P = Q = 2$, and $P = 1, Q = 3$. And we compare these solutions to CP solutions with $R = PQ$.

After rotating, the solution with $P = 1$ and $Q = 2$ is as follows:

$$\mathbf{B} = \begin{pmatrix} 0.68 \\ 0.60 \\ 0.41 \end{pmatrix} \begin{matrix} \text{G} \\ \text{D}, \\ \text{L} \end{matrix} \qquad \mathbf{C} = \begin{pmatrix} 0.01 & \mathbf{0.99} \\ \mathbf{0.70} & -0.09 \\ \mathbf{0.71} & 0.06 \end{pmatrix} \begin{matrix} \text{PN} \\ \text{PC}, \\ \text{SC} \end{matrix}$$

$$\mathbf{G} = \begin{pmatrix} \mathbf{-1.29} & -0.14 \\ 0.12 & \mathbf{1.19} \end{pmatrix}, \quad \mathbf{\Psi} = \begin{pmatrix} \mathbf{1.68} & 0.33 \\ 0.33 & \mathbf{1.44} \end{pmatrix}. \tag{5.7}$$

The unique variance is: diag($\mathbf{U}$) = (0.18 0.38 0.59 0.49 0.47 0.67 0.15 0.55 0.68). The percentage of explained common variance is: 64.41. The percentages of explained common variance for each trait-method combination are given in column 2 of Table 5.4. They are between 41 and 90 percent, where the percentages of L-PC and L-SC are rather low (45% and 41%). Since both matrices $\mathbf{B}$ and $\mathbf{C}$ are rescaled to have columns of length 1, $\mathbf{C}$ is very close to the following form

$$\begin{pmatrix} 0 & 1 \\ * & 0 \\ * & 0 \end{pmatrix},$$

and $\mathbf{\Psi}$ is symmetric, it follows that we obtain 2 parameters from $\mathbf{B}$, 1 parameter from $\mathbf{C}$, and 3 parameters from $\mathbf{\Psi}$. Hence, the model has 6 parameters in total.

For interpretation of (5.7) we also compute $(\mathbf{C}\otimes\mathbf{B})\mathbf{\Psi}$; see columns 2,3 of Table 5.3. The solution (5.7) can be interpreted as follows. Only entries on the diagonal of $\mathbf{\Psi}$ are large and each diagonal entry of $\mathbf{\Psi}$ corresponds to a column of $\mathbf{C}\otimes\mathbf{B} = [\mathbf{c}_1 \otimes \mathbf{B} \mid \mathbf{c}_2 \otimes \mathbf{B}]$. Factor 1 is interpreted as a general factor but only for Peer Checklist ratings (PC) and Supervisor Checklist ratings (SC). Factor 2 is also a general factor but only for Peer Nominations (PN). Both factors have smaller loadings for Ability to Apply Learning (L).

Compared to the solution (5.5) using CP, matrix $\mathbf{C}$ in (5.7) has a clearer interpretation than $\mathbf{C}$ in (5.5). Matrix $\mathbf{B}$ in (5.7) has only one column which is very similar to the first column of $\mathbf{B}$ in (5.5). Since the Tucker3 factor model has less parameters than those of the CP factor model, solution (5.7) is more parsimonious than solution (5.5).

After rotating, the solution with $P = 2$ and $Q = 2$ is as follows:

$$\mathbf{B} = \begin{pmatrix} 0.70 & -0.65 \\ 0.60 & 0.35 \\ 0.38 & 0.66 \end{pmatrix} \begin{matrix} \text{G} \\ \text{D}, \\ \text{L} \end{matrix} \qquad \mathbf{C} = \begin{pmatrix} 0.14 & \mathbf{0.97} \\ \mathbf{0.61} & -0.23 \\ \mathbf{0.78} & 0.00 \end{pmatrix} \begin{matrix} \text{PN} \\ \text{PC}, \\ \text{SC} \end{matrix}$$

$$\mathbf{G} = \begin{pmatrix} 0.08 & 0.03 & \mathbf{1.13} & 0.06 \\ \mathbf{-1.33} & 0.00 & -0.12 & -0.10 \\ -0.01 & -0.59 & -0.01 & -0.37 \\ -0.01 & -0.01 & -0.00 & -0.31 \end{pmatrix}, \quad \mathbf{\Psi} = \begin{pmatrix} \mathbf{1.79} & 0.00 & 0.26 & 0.14 \\ 0.00 & 0.35 & 0.05 & 0.22 \\ 0.26 & 0.05 & \mathbf{1.30} & 0.08 \\ 0.14 & 0.22 & 0.08 & 0.25 \end{pmatrix}. \tag{5.8}$$

The unique variance is: diag($\mathbf{U}$) = (0.30 0.34 0.60 0.50 0.47 0.66 0.00 0.56 0.67). The percentage of explained common variance is: 76.11. The percentages of explained common variance for each trait-method combination are given in column 4 of Table 5.4. They are between 46 and 98 percent, where the percentage of G-PC is rather low (46%). Since both $\mathbf{B}$ and $\mathbf{C}$ are rescaled to have columns

of length 1, $\mathbf{C}$ has one entry that equals to zero, and $\mathbf{\Psi}$ is symmetric with one zero entry, this implies that $\mathbf{B}$ has 4 parameters, $\mathbf{C}$ has 3 parameters, and $\mathbf{\Psi}$ has 9 parameters. Hence, the model has 16 parameters in total.

For interpretation of (5.8) we also compute $(\mathbf{C} \otimes \mathbf{B})\mathbf{\Psi}$; see columns 4, 5, 6, 7 of Table 5.3. The solution (5.8) can be interpreted as follows. Only the first and third entries on the diagonal of $\mathbf{\Psi}$ are large and each diagonal entry of $\mathbf{\Psi}$ corresponds to a column of $\mathbf{C} \otimes \mathbf{B} = [\mathbf{c}_1 \otimes \mathbf{b}_1 \mid \mathbf{c}_1 \otimes \mathbf{b}_2 \mid \mathbf{c}_2 \otimes \mathbf{b}_1 \mid \mathbf{c}_2 \otimes \mathbf{b}_2]$. This implies that columns $\mathbf{c}_1 \otimes \mathbf{b}_2$ and $\mathbf{c}_2 \otimes \mathbf{b}_2$ play only a minor role in the interpretation. Hence, this solution is interpreted almost the same to the interpretation of the solution (5.7).

After rotating, the result with $P = 1$ and $Q = 3$ is:

$$\mathbf{B} = \begin{pmatrix} 0.68 \\ 0.60 \\ 0.42 \end{pmatrix} \begin{matrix} \text{G} \\ \text{D}, \\ \text{L} \end{matrix} \qquad \mathbf{C} = \begin{pmatrix} \mathbf{1.00} & -0.01 & -0.01 \\ 0.01 & 0.01 & \mathbf{1.00} \\ 0.01 & \mathbf{1.00} & -0.01 \end{pmatrix} \begin{matrix} \text{PN} \\ \text{PC}, \\ \text{SC} \end{matrix}$$

$$\mathbf{G} = \begin{pmatrix} \mathbf{1.18} & 0.12 & 0.07 \\ -0.06 & -0.16 & \mathbf{-1.07} \\ -0.11 & \mathbf{-1.13} & -0.18 \end{pmatrix}, \quad \mathbf{\Psi} = \begin{pmatrix} \mathbf{1.42} & 0.29 & 0.17 \\ 0.29 & \mathbf{1.33} & 0.40 \\ 0.17 & 0.40 & \mathbf{1.19} \end{pmatrix}. \tag{5.9}$$

The unique variance is: diag($\mathbf{U}$) = (0.24 0.36 0.60 0.49 0.47 0.67 0.08 0.56 0.67). The percentage of explained common variance is: 81.69. The explained common variances for all trait-method combinations are given in column 6 of Table 5.4. They are between 58 and 91 percent. Most trait-method combinations with low percentages of explained common variance in the solutions with $P = 1, Q = 2$ and $P = Q = 2$ have much larger percentage of explained common variance in the $P = 1, Q = 3$ solution. Since both $\mathbf{B}$ and $\mathbf{C}$ are rescaled to have columns of length 1, $\mathbf{C}$ is almost completely fixed, and $\mathbf{\Psi}$ is symmetric, this follows that we obtain 2 parameters from $\mathbf{B}$, and 6 parameters from $\mathbf{\Psi}$. Therefore, the model

has 8 parameters in total.

For interpretation of (5.9) we also compute $(\mathbf{C} \otimes \mathbf{B})\mathbf{\Psi}$; see columns 8, 9, 10 of Table 5.3. The solution (5.9) can be interpreted as follows. Since only entries on the diagonal of $\mathbf{\Psi}$ are large and each diagonal entry of $\mathbf{\Psi}$ corresponds to a column of $\mathbf{C} \otimes \mathbf{B} = [\mathbf{c}_1 \otimes \mathbf{B} \mid \mathbf{c}_2 \otimes \mathbf{B} \mid \mathbf{c}_3 \otimes \mathbf{B}]$, this implies that we obtain one method per factor. Factor 1 is a general factor but only for Supervisor Checklist ratings (SC). Factor 2 is a general factor but only for Peer Nominations (PN). Factor 3 is a general factor but only for Peer Checklist ratings (PC). All factors having smaller loadings for Ability to Apply Learning (L).

Compared to the solutions (5.5) and (5.6) using CP, matrices $\mathbf{C}$ in the Tucker3 solutions (5.7) and (5.9) have much more weights that are very close to zero. And each column of $\mathbf{C}$ in Tucker3 solution has at least one entry close to zero and at least one entry close to 1. This means that matrices $\mathbf{C}$ in Tucker3 solutions have a clearer interpretation than $\mathbf{C}$ in CP solutions. Matrices $\mathbf{B}$ in (5.7) and (5.9) have only one column instead of two similar columns of $\mathbf{B}$ in (5.5) and (5.6).

Since Tucker3 solutions have less parameters than the CP solutions with $R = PQ$ factors, the Tucker3 solutions are more parsimonious than the corresponding CP solutions. However, the percentages of explained common variance of Tucker3 are a little bit less than of CP solutions with $R = PQ$.

Table 5.3: Values of $(\mathbf{C} \otimes \mathbf{B})\,\mathbf{\Psi}$ for the estimated Tucker3 covariance model (5.2) with P traits factors and Q method factors, fitted to the MTMM data from Dickinson and Tice (1973).

variable	$P=1,Q=2$		$P=2,Q=2$				$P=1,Q=3$		
G-PN	0.24	**0.98**	0.26	-0.14	**0.86**	-0.11	**0.97**	0.18	0.10
D-PN	0.22	**0.87**	0.35	0.12	**0.81**	0.16	**0.85**	0.16	0.09
L-PN	0.14	**0.59**	0.29	0.19	**0.56**	0.22	**0.59**	0.11	0.06
G-PC	**0.78**	0.06	**0.75**	-0.11	-0.10	-0.00	0.13	0.29	**0.82**
D-PC	**0.69**	0.05	**0.61**	0.05	-0.07	0.06	0.12	0.25	**0.72**
L-PC	**0.47**	0.03	0.37	0.10	-0.04	0.07	0.08	0.18	**0.50**
G-SC	**0.83**	0.22	**0.98**	-0.18	0.12	-0.03	0.21	**0.91**	0.26
D-SC	**0.74**	0.20	**0.84**	0.09	0.14	0.13	0.18	**0.79**	0.22
L-SC	**0.50**	0.13	**0.54**	0.18	0.10	0.16	0.12	**0.56**	0.15

Note: Numbers larger than 0.4 are in boldfont.

Table 5.4: The explained common variances for all trait-method combinations and the communalities for the MTMM data from Dickinson and Tice (1973)

variable	$P = 1, Q = 2$		$P = 2, Q = 2$		$P = 1, Q = 3$		CP with $R = 2$		CP with $R = 3$	
	ECV%	comm.	ECV%	comm.	ECV%	comm.	ECV%	comm.	ECV%	comm.
G-PN	78	0.82	98	0.70	81	0.75	79	0.82	78	0.75
D-PN	90	0.61	93	0.66	91	0.63	91	0.62	94	0.63
L-PN	58	0.40	53	0.39	58	0.40	60	0.41	65	0.40
G-PC	58	0.51	46	0.49	91	0.51	55	0.51	90	0.51
D-PC	65	0.53	66	0.52	90	0.53	62	0.53	90	0.53
L-PC	45	0.33	51	0.33	82	0.33	44	0.33	85	0.33
G-SC	61	0.85	81	1.00	85	0.91	62	0.85	88	0.91
D-SC	55	0.44	60	0.43	81	0.44	57	0.45	80	0.43
L-SC	41	0.32	70	0.33	58	0.32	45	0.32	59	0.33
ECV% total	64.41		76.11		81.69		64.68		82.76	

5.4 Application II

In this section, we apply our three-mode Tucker3 factor model (5.2) to BJW data from section 4.3.2 for which we fitted three-mode CP factor model (4.9). To compare to the CP solutions in section 4.3.2, we fit three-mode Tucker3 factor model (5.2) to this dataset with $P \in \{2, 3\}$ and $Q \in \{1, 2\}$. The explained common variances for $P \in \{2, 3\}$, and $Q \in \{1, 2\}$ are given in the following table.

P	Q	ECV% for Tucker3	ECV% for CP with $R = PQ$
2	1	71.00	74.51
3	1	74.76	84.54
2	2	86.00	
3	2	91.52	

Below, we present the solutions with $P = 2, Q = 1$ and $P = 3, Q = 1$.

After rotating, the result with $P = 2$ and $Q = 1$ is:

$$\mathbf{B} = \begin{pmatrix} 0.33 & 0.13 \\ 0.41 & -0.17 \\ 0.41 & -0.25 \\ 0.33 & 0.28 \\ 0.10 & \mathbf{0.78} \\ 0.33 & 0.31 \\ 0.38 & -0.25 \\ 0.40 & -0.14 \end{pmatrix} \begin{matrix} \text{item 1} \\ \text{item 2} \\ \text{item 3} \\ \text{item 4} \\ \text{item 5} \\ \text{item 6} \\ \text{item 7} \\ \text{item 8} \end{matrix}, \qquad \mathbf{C} = \begin{pmatrix} 0.70 \\ 0.72 \end{pmatrix} \begin{matrix} \text{Yourself} \\ \text{Others} \end{matrix},$$

$$\mathbf{G} = \begin{pmatrix} \mathbf{-2.76} & -0.42 \\ 0.07 & \mathbf{1.24} \end{pmatrix}, \quad \mathbf{\Psi} = \begin{pmatrix} \mathbf{7.61} & 1.26 \\ 1.26 & \mathbf{1.73} \end{pmatrix}. \tag{5.10}$$

The unique variance is: diag($\mathbf{U}$) = (0.34 0.12 0.16 0.19 0.30 0.15 0.21 0.11 0.21 0.24 0.15 0.23 0.00 0.11 0.20 0.12). The percentage of explained common variance is: 71.00. The explained common variances for all item-condition combinations are given in column 2 of Table 5.6. They are between 58 and 85 percent with low percentage for items 7 and 8 for Others (58%).

For interpreting (5.10) we compute $(\mathbf{C} \otimes \mathbf{B})\mathbf{\Psi}$; see columns 2, 3 of Table 5.10. The solution (5.10) can be interpreted as follows. Factor 1 is a general factor. Factor 2 is a combination of items 4,5, and 6. There is not much difference between two conditions.

Compared to the CP solution with $R = PQ = 2$, this solution has less percent explained common variance than of the CP solution (71.00% versus 74.51%). In CP solution, there is one general factor and one "for Yourself" factor.

After rotating, the result with $P = 3$ and $Q = 1$ is:

$$\mathbf{B} = \begin{pmatrix} 0.32 & -0.17 & \mathbf{0.72} \\ 0.40 & -0.25 & 0.09 \\ 0.39 & -0.17 & -0.35 \\ 0.35 & 0.17 & 0.24 \\ 0.18 & \mathbf{0.87} & -0.08 \\ 0.36 & 0.23 & 0.17 \\ 0.36 & -0.17 & -0.33 \\ 0.40 & -0.04 & -0.36 \end{pmatrix} \begin{matrix} \text{item 1} \\ \text{item 2} \\ \text{item 3} \\ \text{item 4} \\ \text{item 5} \\ \text{item 6} \\ \text{item 7} \\ \text{item 8} \end{matrix}, \qquad \mathbf{C} = \begin{pmatrix} 0.69 \\ 0.72 \end{pmatrix} \begin{matrix} \text{Yourself} \\ \text{Others} \end{matrix},$$

$$\mathbf{G} = \begin{pmatrix} \mathbf{-2.79} & -0.20 & -0.14 \\ 0.01 & 0.08 & \mathbf{0.76} \\ 0.02 & \mathbf{1.16} & 0.24 \end{pmatrix}, \quad \mathbf{\Psi} = \begin{pmatrix} \mathbf{7.79} & 0.61 & 0.40 \\ 0.61 & \mathbf{1.39} & 0.37 \\ 0.40 & 0.37 & \mathbf{0.64} \end{pmatrix}. \qquad (5.11)$$

The unique variance is: diag($\mathbf{U}$) = (0.34 0.13 0.16 0.19 0.29 0.15 0.21 0.11 0.21 0.25 0.15 0.23 0.00 0.11 0.19 0.11). The percentage of explained common variance

is: 74.76. The explained common variances for all item-condition combinations are given in column 4 of Table 5.6. They are between 62 and 89 percent.
For interpreting (5.10) we compute $(\mathbf{C} \otimes \mathbf{B})\mathbf{\Psi}$; see columns 4, 5, 6 of Table 5.10. The solution (5.11) can be interpreted as follows. Factor 1 represents a general factor. Factor 2 represents item 5. Factor 3 represents mostly item 1. There is not much different between two conditions.

In the CP solution with $R = PQ = 3$ factors (see Table (4.4)), factor 1 is a combination of items 2,3,7,8 and has higher loadings for Yourself; factor 2 is a combination of items 1,4,5,6 and has higher loadings for Others; factor 3 is a contrast factor between Yourself and Others. This implies that solution (5.11) is not clearer than the CP solution. Additionally, solution (5.11) has less percent explained common variance than of the CP solution (74.76% versus 84.54%).

Clearly, to estimate the BJW data set, the three-mode CP factor model (4.9) is more useful than the three-mode Tucker3 factor model (5.2). Whereas, for the data set from Dickinson and Tice (1973), the three-mode Tucker3 factor model (5.2) is more useful. Hence, it depends on the data set, which model has a clearer interpretation.

Table 5.5: Values of $(\mathbf{C} \otimes \mathbf{B})\, \mathbf{\Psi}$ for the estimated Tucker3 covariance model (5.2) with $P = 1$, $Q \in \{2, 3\}$ factors fitted to the BJW data from section 4.3.2.

variable	$P = 2, Q = 1$		$P = 3, Q = 1$		
item 1, yourself	**1.88**	0.45	**1.90**	0.15	0.37
item 2, yourself	**2.06**	0.16	**2.05**	-0.05	0.08
item 3, yourself	**1.94**	0.05	**1.94**	-0.09	-0.09
item 4, yourself	**2.00**	**0.64**	**2.04**	0.38	0.25
item 5, yourself	**1.26**	**1.04**	**1.34**	**0.89**	0.23
item 6, yourself	**2.06**	**0.68**	**2.10**	0.42	0.24
item 7, yourself	**1.84**	0.03	**1.83**	-0.09	-0.09
item 8, yourself	**2.01**	0.18	**2.02**	0.02	-0.06
item 1, others	**1.94**	0.47	**1.98**	0.16	0.38
item 2, others	**2.12**	0.17	**2.13**	-0.06	0.09
item 3, others	**2.01**	0.05	**2.01**	-0.09	-0.09
item 4, others	**2.06**	**0.66**	**2.13**	0.40	0.26
item 5, others	**1.30**	**1.07**	**1.40**	**0.93**	0.24
item 6, others	**2.13**	**0.70**	**2.19**	0.44	0.25
item 7, others	**1.89**	0.04	**1.90**	-0.10	-0.09
item 8, others	**2.08**	0.19	**2.10**	0.03	-0.06

Note: Numbers larger than 0.5 are in boldfont.

Table 5.6: The explained common variances for all item-condition combinations and the communalities for the BJW data from section 4.3.2

variable	$P = 2, Q = 1$		$P = 3, Q = 1$		CP with $R = 2$		CP with $R = 3$	
	ECV%	comm.	ECV%	comm.	ECV%	comm.	ECV%	comm.
item 1, yourself	68	0.66	87	0.66	69	0.66	70	0.66
item 2, yourself	70	0.88	71	0.87	86	0.88	85	0.87
item 3, yourself	68	0.84	67	0.84	93	0.84	94	0.84
item 4, yourself	70	0.81	71	0.81	74	0.81	82	0.81
item 5, yourself	79	0.70	86	0.71	43	0.70	79	0.71
item 6, yourself	72	0.85	72	0.85	68	0.85	81	0.85
item 7, yourself	68	0.79	89	0.79	85	0.79	88	0.79
item 8, yourself	70	0.89	72	0.89	89	0.89	89	0.89
item 1, others	68	0.79	85	0.79	74	0.79	72	0.79
item 2, others	76	0.76	78	0.75	76	0.76	83	0.75
item 3, others	66	0.85	70	0.85	65	0.85	88	0.85
item 4, others	85	0.77	85	0.77	89	0.77	90	0.77
item 5, others	81	1.00	85	1.00	62	1.00	85	1.00
item 6, others	78	0.89	79	0.89	86	0.89	86	0.89
item 7, others	58	0.80	62	0.81	62	0.81	90	0.81
item 8, others	58	0.88	62	0.89	66	0.88	86	0.89
ECV% total	71.00		74.76		74.51		84.54	

5.5 Discussion

In this chapter, we have proposed and demonstrated a method for three-mode factor analysis using MRFA to estimate unique variances $\mathbf{U}$ and Tucker3 to estimate the covariance matrix of the common part. This method is an extension of the CP-based method that is proposed in chapter 4.

By using the Tucker3 covariance model, the component matrices and the core array of interactions of the factors in the second and the third modes are not unique. Therefore we use the Joint Othomax rotation of (Kiers, 1998a) to rotate the Tucker3 solution to simple structure in matrices corresponding to the second and the third modes, and also in the core array.

The solutions in section 5.3 show that the Tucker3 solutions have a clearer interpretation than the CP solutions for the MTMM data of Dickinson and Tice (1973). However, the solutions in section 5.4 show that the CP solutions have a clearer interpretation than the Tucker3 solutions for the BJW data of section 4.3.2. Hence, the Tucker3 factor model is not always more appropriate than the CP factor model and vice versa. Therefore, we advise to use both models on the same dataset and to choose the model with the clearest interpretation (provided the explained common variances are similar in size).

Chapter 6

Multi-set factor analysis by means of Parafac2

Abstract

We consider multi-set data consisting of N_k observations, $k = 1, \ldots, K$, (e.g., subject scores) on J variables in K different samples. We introduce a factor model for the $J \times J$ covariance matrices $\mathbf{\Sigma}_k$, $k = 1, \ldots, K$, where the common part is modelled by Parafac2 and the unique variance matrices $\mathbf{U}_k$ are diagonal, $k = 1, \ldots, K$. The Parafac2 model implies a common loadings matrix that is rescaled for each k, and a common factor correlation matrix. We estimate the unique variances $\mathbf{U}_k$ by Minimum Rank Factor Analysis on $\mathbf{\Sigma}_k$ for each k. The factors can be chosen orthogonal or oblique. We present a novel algorithm to estimate the Parafac2 part and demonstrate its performance in a simulation study. Also, we fit our model to a dataset in the literature. Our model is easy to estimate and interpret. The unique variances, the factor correlation matrix and the communalities are guaranteed to be proper, and a percentage of explained common variance can be computed for each k. Also, the Parafac2 part is rotationally unique under mild conditions.

This chapter is based on: Lam, T.T.T. and Stegeman, A. (2014). Multi-set factor analysis by means of Parafac2. Submitted.

6.1 Introduction

We consider exploratory factor analysis of multi-set data. Data are called *multi-set* when they consist of a set of data matrices and each data matrix has the same column units (i.e., variables) but different row units (i.e., observational units). In particular, we consider multi-set data in which the same variables are observed for several different populations or subpopulations. Under the assumption that the same latent factors underly the observed data in each (sub)population, exploratory factor analysis can be used to estimate factor loadings and the strengths of the factors in each (sub)population. The fit of the factor model to each data matrix can then be used to assess the hypothesis that the same latent factors are underlying the observed data in each (sub)population.

Next, we formally introduce our model. Let $\mathbf{X}_k$ be the data matrix of the sample from (sub)population k, for $k = 1, \ldots, K$. We measure the same J variables in each sample, where we have N_k observations in sample k. Hence, matrix $\mathbf{X}_k$ has J columns and N_k rows, $k = 1, \ldots, K$. We assume that the columns of each $\mathbf{X}_k$ are centered. For R underlying factors, our exploratory factor model is

$$\mathbf{X}_k = \mathbf{F}_k \mathbf{C}_k \mathbf{B}^T + \mathbf{E}_k, \quad k = 1, \ldots, K, \tag{6.1}$$

where $\mathbf{F}_k$ $(N_k \times R)$ is the matrix of factor scores in sample k, matrix $\mathbf{B}$ $(J \times R)$ is a loading matrix common to all samples, $\mathbf{C}_k$ $(R \times R)$ is a diagonal matrix containing the factor strengths in sample k, and $\mathbf{E}_k$ $(N_k \times J)$ is the unique part of sample k. The common (correlation producing) part of sample k is thus modeled as $\mathbf{F}_k \mathbf{C}_k \mathbf{B}^T$. It is assumed that $N_k^{-1} \mathbf{F}_k^T \mathbf{F}_k = \mathbf{\Phi}$ is constant over k. Together with the constant loading matrix $\mathbf{B}$, this expresses the hypothesis of the same factors underlying the observations in all samples. The loading matrix of sample k can be written as $\mathbf{B}_k = \mathbf{B}\mathbf{C}_k$. Hence, $\mathbf{B}_k$ is congruent to each $\mathbf{B}_l$

for $k \neq l$. As is usual in factor analysis, the unique part of variable j in sample k is assumed to be uncorrelated to the unique part of other variables in the same sample k, and also to the factors in sample k. When $\mathbf{E}_k$ are considered as residuals for sample k, model (6.1) becomes a multi-set component model and is known as direct Parafac2 (Harshman, 1972).

As is usual in factor analysis, we estimate our multi-set exploratory factor model on the covariance or correlation matrices of each sample k. Under the assumptions of uncorrelatedness above, we rewrite (6.1) for the covariance matrices $\mathbf{\Sigma}_k = N_k^{-1}\mathbf{X}_k^T\mathbf{X}_k$, $k = 1, \ldots, K$, as follows:

$$\mathbf{\Sigma}_k = \mathbf{B}\mathbf{C}_k\mathbf{\Phi}\mathbf{C}_k\mathbf{B}^T + \mathbf{U}_k, \quad k = 1, \ldots, K, \tag{6.2}$$

where $\mathbf{U}_k$ $(J \times J)$ is a diagonal matrix containing the unique variances for sample k. Throughout, the factors $\mathbf{F}_k$ are scaled to variance one. Hence, $\mathbf{\Phi} = N_k^{-1}\mathbf{F}_k^T\mathbf{F}_k$ can be interpreted as the factor correlation matrix. The factor model (6.2) reveals a second interpretation of the factor strengths in the diagonal matrices $\mathbf{C}_k$. Namely, matrix $\mathbf{C}_k\mathbf{\Phi}\mathbf{C}_k$ can be seen as the factor covariance matrix in sample k. Hence, $\mathbf{C}_k$ contains the factor standard deviations for sample k. When $\mathbf{U}_k$ are replaced by residual matrices, model (6.2) is known as indirect Parafac2 (Harshman, 1972). We refer to (6.2) as the multi-set Parafac2 factor model.

In the estimation procedure for the multi-set Parafac2 factor model (6.2), we first estimate the unique variances $\mathbf{U}_k$ by Minimum Rank Factor Analysis on each $\mathbf{\Sigma}_k$ separately. This guarantees that $\mathbf{\Sigma}_k - \mathbf{U}_k$ is a covariance matrix (i.e., positive semi-definite) for each k. Next, we estimate $\mathbf{B}$, $\mathbf{C}_k$, and $\mathbf{\Phi}$ by fitting the indirect Parafac2 model to $\mathbf{\Sigma}_k - \mathbf{U}_k$, $k = 1, \ldots, K$. For this, we provide a novel indirect Parafac2 algorithm which works as follows. Matrices $\mathbf{Y}_k$ are obtained from the eigendecomposition of $\mathbf{\Sigma}_k - \mathbf{U}_k$ such that $\mathbf{\Sigma}_k - \mathbf{U}_k = \mathbf{Y}_k^T\mathbf{Y}_k$. Then the

direct Parafac2 model (6.1) is fitted to $\mathbf{Y}_k^T$, $k = 1, \ldots, K$, by using the alternating least squares algorithm of Kiers et al. (1999). Since the $\mathbf{\Sigma}_k - \mathbf{U}_k$ are covariance matrices, we can compute the explained common variance for each variable and each sample separately. In our estimation procedure, we can specify whether the factors should be orthogonal or oblique. In case of orthogonal factors, we have $\mathbf{\Phi} = \mathbf{I}_R$ and the covariance model for the common part becomes $\mathbf{B}\mathbf{C}_k^2\mathbf{B}^T$, $k = 1, \ldots, K$, which is also known as the Indscal model (Carroll & Chang, 1970).

In exploratory factor or component analysis of one data matrix, the factors or components can be rotated. A rotation may be found to obtain a loading matrix that is easy to interpret or to obtain a loading matrix that is close to some target matrix. In our multi-set Parafac2 factor model (6.2), the solution $(\mathbf{B}, \mathbf{C}_1, \ldots, \mathbf{C}_K, \mathbf{\Phi})$ of the common part is rotationally unique under some conditions. Next, we present a formal definition of such uniqueness.

Definition 6.1.1 (Kiers et al., 1999)
An indirect Parafac2 solution $(\mathbf{B}, \mathbf{C}_1, \ldots, \mathbf{C}_K, \mathbf{\Phi})$ is called "essentially unique" if for any other solution $(\bar{\mathbf{B}}, \bar{\mathbf{C}}_1, \ldots, \bar{\mathbf{C}}_K, \bar{\mathbf{\Phi}})$ with $\mathbf{B}\mathbf{C}_k\mathbf{\Phi}\mathbf{C}_k\mathbf{B}^T = \bar{\mathbf{B}}\bar{\mathbf{C}}_k\bar{\mathbf{\Phi}}\bar{\mathbf{C}}_k\bar{\mathbf{B}}^T$, $k = 1, \ldots, K$, there exists a permutation matrix $\mathbf{P}$ and diagonal scaling matrices $\mathbf{T}_b$, $\mathbf{T}_c$, $\mathbf{T}_{\mathbf{\Phi}}$ with $\mathbf{T}_b\mathbf{T}_c\mathbf{T}_{\mathbf{\Phi}} = \mathbf{I}_R$ such that

(i) $\bar{\mathbf{B}} = \mathbf{B}\mathbf{P}\mathbf{T}_b$,

(ii) $\bar{\mathbf{C}}_k = \lambda_k\mathbf{P}^T\mathbf{C}_k\mathbf{P}\mathbf{T}_c$, where $\lambda_k = \pm 1$, $k = 1 \ldots K$,

(iii) $\bar{\mathbf{\Phi}} = \mathbf{T}_{\mathbf{\Phi}}\mathbf{P}^T\mathbf{\Phi}\mathbf{P}\mathbf{T}_{\mathbf{\Phi}}$.

Throughout, we scale Parafac2 solutions such that $\mathbf{\Phi}$ has a diagonal of ones, and $\mathbf{B}$ has columns with sum of squares unity.

For the case of orthogonal factors ($\mathbf{\Phi} = \mathbf{I}_R$), from condition (iii) we have $(\mathbf{T}_{\mathbf{\Phi}})^2 = \mathbf{I}_R$ and $\bar{\mathbf{\Phi}} = \mathbf{I}_R$, and

$$\bar{\mathbf{B}}(\bar{\mathbf{C}}_k)^2\bar{\mathbf{B}}^T = \mathbf{B}(\mathbf{C}_k)^2\mathbf{B}^T, \quad k = 1 \ldots K. \tag{6.3}$$

Hence, the entries of $\mathbf{C}_k$ are identified up to sign only when $\mathbf{\Phi} = \mathbf{I}_R$.

Ten Berge and Kiers (1996) prove the following sufficient condition for uniqueness of indirect Parafac2 with $R = 2$: at least four covariance matrices are analyzed (i.e., $K \geq 4$), $\mathbf{\Phi}$ is positive definite, rank($\mathbf{B}$) $= R$, and at least four $\mathbf{C}_k$ matrices are nonsingular and mutually not proportional. These results sharpen the results by Harshman and Lundy (1996). Due to these results, indirect Parafac2 with $R = 2$ is usually unique with four matrices; with three matrices it is not unique unless a certain additional assumption is introduced. If the matrices $\mathbf{C}_k$ are constrained to be nonnegative, three matrices are enough to have uniqueness for $R = 2$. Kiers et al. (1999) prove the equivalence of uniqueness in direct Parafac2 to uniqueness in indirect Parafac2 under mild conditions. Based on simulations, Kiers et al. (1999) suggest that Parafac2 is usually unique for $K \geq 4$ and any $R \geq 2$.

Our exploratory multi-set Parafac2 factor model (6.2) is the factor analogue of the simultaneous component models SCA-PF2 (for oblique components and direct Parafac2 form) and SCA-IND (for orthogonal components and Indscal form) presented in Timmerman and Kiers (2003). The difference between our factor model (6.2) and the component models SCA-PF2 and SCA-IND is the same as the difference between exploratory common factor analysis (EFA) and principal component analysis (PCA) for one data matrix. That is, our factor model explicitly takes into account the unique part of each variable while this is neglected in the component models. The unique part may be seen as measurement error or as a part of the variable that does not correlate with other variables

or as a combination of both. The lively discussion of EFA versus PCA has been documented extensively in the literature; see e.g. Velicer and Jackson (1990) for an overview. PCA can be favored because of its computational simplicity and manifest component approach, whereas EFA is computationally more difficult and features latent factors. PCA and common EFA are said to yield similar estimated loadings in simulation studies, although this is not generally true and theoretical results prove only asymptotical similarity (e.g., Ogasawara, 2000) or for the case of equal unique variances (e.g., Tipping & Bishop, 1999). We believe that whenever measurement error (or a unique part in general) should be taken into account explicitly, a factor analytic model should be applied. Naturally, this also applies to multi-set data.

A class of methods that also considers multi-set factor analysis are the multigroup factor analysis models originating with Jöreskog (1971a) and Sörbom (1974). These models are usually used in a confirmatory way. In multigroup confirmatory factor analysis (CFA), or single-group CFA methods, a prespecified factor pattern can be tested by means of a chi-square test. That is, zero restrictions on the loading matrix $\mathbf{B}$ can be tested. In multigroup CFA also hypotheses such as equal loadings per group, equal unique variances, or equal factor means can be tested to assess measurement invariance over the groups. Our exploratory multi-set Parafac2 factor model (6.2) relates to multigroup CFA models as EFA relates to CFA for a single group. The outcomes of an analysis by our multi-set model (6.2) can serve as input for a multigroup CFA approach. That is, from an exploratory multi-set factor analysis several hypothetical factor patterns can be inferred which can then be tested in the multigroup CFA setting.

In terms of methodology, there are two crucial differences between our exploratory multi-set Parafac2 factor model (6.2) and existing (exploratory or confirmatory) multigroup factor analysis models based on Jöreskog (1971a) and

Sörbom (1974). First, our model does not require any distributional assumptions, while these are present in the use of existing models. For example, normality assumptions should be checked when using chi-squared tests to compare nested models, and deviations from normality should be taken into account. Second, for the existing models $\mathbf{\Sigma}_k - \mathbf{U}_k$, $k = 1, \ldots, K$, are not necessarily covariance matrices. Hence, a percentage of explained common variance cannot be computed, which is essential in determining the fit for the common part of the data for each (sub)population k. As an example, consider the analysis of a multigroup dataset of $J = 9$ tests administered to $K = 4$ groups of school children, taken from Holzinger and Swineford (1939). The dataset is analyzed by Jöreskog (1971a) and Sörbom (1974) using nearly the same multigroup factor model. In the solutions of both Jöreskog (1971a) and Sörbom (1974) the matrices $\mathbf{\Sigma}_k - \mathbf{U}_k$ have negative and positive eigenvalues for all k. Hence, they are not covariance matrices.

Our model and estimation procedure are related to the exploratory three-mode factor model of Stegeman and Lam (2014). In three-mode data, matrix $\mathbf{X}_k$ contains N observations of J variables for occassion or condition k, for $k = 1, \ldots, K$. Contrary to Parafac2, the observational units in the rows of each $\mathbf{X}_k$ are assumed to be the same. Besides covariance matrices $\mathbf{\Sigma}_k = N_k^{-1}\mathbf{X}_k^T\mathbf{X}_k$, we can also compute $\mathbf{\Sigma}_{kl} = N_k^{-1}\mathbf{X}_k^T\mathbf{X}_l$ in which covariances between the variables for conditions k and l are given. The three-mode factor model of Stegeman and Lam (2014) for the complete $JK \times JK$ covariance matrix $\mathbf{\Sigma}$ is given by $\mathbf{\Sigma} = (\mathbf{C} \odot \mathbf{B})\mathbf{\Phi}(\mathbf{C} \odot \mathbf{B}) + \mathbf{U}$, which can be written as

$$\begin{bmatrix} \mathbf{\Sigma}_{11} & \cdots & \mathbf{\Sigma}_{1K} \\ \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{K1} & \cdots & \mathbf{\Sigma}_{KK} \end{bmatrix} = \begin{pmatrix} \mathbf{BC}_1 \\ \vdots \\ \mathbf{BC}_K \end{pmatrix} \mathbf{\Phi} \begin{pmatrix} \mathbf{BC}_1 \\ \vdots \\ \mathbf{BC}_K \end{pmatrix}^T + \begin{bmatrix} \mathbf{U}_1 & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & \mathbf{U}_K \end{bmatrix}. \quad (6.4)$$

For multi-set data, the samples have different observational units and sample sizes. Hence, we do not have the covariance matrices $\mathbf{\Sigma}_{kl}$ for $k \neq l$. In this case, only the diagonal blocks in (6.4) remain. For diagonal block k we obtain: $\mathbf{\Sigma}_{kk} = \mathbf{BC}_k\mathbf{\Phi}\mathbf{C}_k\mathbf{B}^T + \mathbf{U}_k$, which corresponds to our multi-set factor model (6.2). In the estimation procedure of Stegeman and Lam (2014) for the three-mode factor model (6.4), first the diagonal matrix $\mathbf{U}$ of unique variances is computed via Minimum Rank Factor Analysis. Next, $\mathbf{\Sigma} - \mathbf{U}$ is decomposed as $\mathbf{\Sigma} - \mathbf{U} = \mathbf{Y}^T\mathbf{Y}$ and $\mathbf{Y} \approx \mathbf{F}(\mathbf{C} \odot \mathbf{B})^T$ is fitted. The latter is the Candecomp/Parafac model (Carroll & Chang, 1970; Harshman, 1970). Conceptually, this shows the similarity between our multi-set Parafac2 factor model (6.2) and its estimation procedure, and the three-mode Candecomp/Parafac factor model (6.4) and estimation procedure of Stegeman and Lam (2014).

It may be instructive to list the objective functions of indirect and direct Parafac2 related to our multi-set Parafac2 factor model (6.2) in the following table.

	Parafac2	Parafac2 FA
indirect	(A) $\sum_k \|\|\mathbf{\Sigma}_k - \mathbf{BC}_k\mathbf{\Phi}\mathbf{C}_k\mathbf{B}^T\|\|^2$	(C) $\sum_k \|\|(\mathbf{\Sigma}_k - \mathbf{U}_k) - \mathbf{BC}_k\mathbf{\Phi}\mathbf{C}_k\mathbf{B}^T\|\|^2$
direct	(B) $\sum_k \|\|\tilde{\mathbf{Y}}_k - \mathbf{F}_k\mathbf{C}_k\mathbf{B}^T\|\|^2$	(D) $\sum_k \|\|\mathbf{Y}_k - \mathbf{F}_k\mathbf{C}_k\mathbf{B}^T\|\|^2$

Here, $||\mathbf{Z}||$ stands for the Frobenius norm of matrix $\mathbf{Z}$, which is defined as the square root of the sum of squares of all entries of $\mathbf{Z}$. Model (A) is indirect Parafac2, which can be fitted by direct Parafac2 in (B) with $\tilde{\mathbf{Y}}_k^T\tilde{\mathbf{Y}}_k = \mathbf{\Sigma}_k$, for $k = 1, \ldots, K$. Model (C) is our multi-set Parafac2 factor model in indirect form, which can be fitted (given $\mathbf{U}_k$) by direct Parafac2 in (D) with $\mathbf{Y}_k^T\mathbf{Y}_k = \mathbf{\Sigma}_k - \mathbf{U}_k$,

for $k = 1, \ldots, K$. To describe the difference in fitting the models, (A) and (C) could be named *covariance fitting*, and (B) and (D) could be named *decomposed covariance fitting*.

As mentioned in Kiers et al. (1999), the estimation procedure of the indirect Parafac2 model (A) that is given in Kiers (1993) is computationally complex and inefficient, whereas the estimation procedure of the direct Parafac2 model (B) as given in Kiers et al. (1999) is simpler and more efficient. Therefore, we fit the indirect Parafac2 model (A) by using direct Parafac2 in (B) with $\tilde{\mathbf{Y}}_k^T \tilde{\mathbf{Y}}_k = \mathbf{\Sigma}_k$. Apart from the estimation of the unique variances $\mathbf{U}_k$, the same reasoning applies to (C) and (D).

This chapter is organized as follows. In section 6.2, we present our estimation procedure (A)-(B) for the indirect Parafac2 model and compare its performance to the indirect Parafac2 algorithm (A) of Kiers (1993). In section 6.3, we present our algorithm (C)-(D) for finding a best solution of the multi-set Parafac2 factor model (6.2). Also, we show how the percentage of explained common variance can be obtained for each variable and each sample k. In section 6.4, we assess the performance of our estimation procedure (C)-(D) in a simulation study. In section 6.5, we apply our multi-set Parafac2 factor model to a dataset in the literature. Finally, section 6.6 contains a discussion of our findings.

6.2 Estimating indirect Parafac2 via direct Parafac2

Here, we present and evaluate our estimation procedure (A)-(B) for indirect Parafac2. In section 6.2.1, we discuss in detail the algorithm of Kiers et al. (1999) for direct Parafac2 (B). In section 6.2.2, we present our estimation procedure for indirect Parafac2 which makes use of the direct Parafac2 algorithm. In section 6.2.3, we conduct a simulation study to compare the performance of our indirect

Parafac2 algorithm to that of Kiers (1993).

6.2.1 Algorithm for direct Parafac2 of Kiers et al. (1999)

Kiers et al. (1999) provided an algorithm to fit the direct Parafac2 model (6.1) by minimizing

$$\sum_{k=1}^{K} \|\mathbf{X}_k - \mathbf{F}_k\mathbf{C}_k\mathbf{B}^T\|^2, \tag{6.5}$$

over $\mathbf{B}$, $\mathbf{F}_k$, and $\mathbf{C}_k$, $k = 1 \dots K$, under the constraint that $\mathbf{F}_k^T\mathbf{F}_k = \mathbf{F}_l^T\mathbf{F}_l$ for all $k \neq l$.

Because of the constraint on $\mathbf{F}_k$, $k = 1, \dots, K$, there exists an $R \times R$ matrix $\mathbf{F}$ and columnwise orthonormal matrices $\mathbf{P}_1, \dots, \mathbf{P}_K$ ($N_k \times R$) such that $\mathbf{F}_k = \mathbf{P}_k\mathbf{F}$, $k = 1, \dots, K$. Hence minimizing (6.5) is equivalent to minimizing

$$\sum_{k=1}^{K} \|\mathbf{X}_k - \mathbf{P}_k\mathbf{F}\mathbf{C}_k\mathbf{B}^T\|^2, \tag{6.6}$$

subject to the constraint that $\mathbf{P}_k^T\mathbf{P}_k = \mathbf{I}_R$, $k = 1, \dots, K$.

To minimize function (6.6), Kiers et al. (1999) proposed an alternating least squares (ALS) algorithm that alternately minimizes (6.6) over $\mathbf{P}_k$ for fixed $\mathbf{F}$, $\mathbf{C}_1, \dots, \mathbf{C}_K$ and $\mathbf{B}$, for $k = 1, \dots, K$, and over $\mathbf{F}$, $\mathbf{C}_1, \dots, \mathbf{C}_K$ and $\mathbf{B}$ for fixed $\mathbf{P}_1, \dots, \mathbf{P}_K$. The two main steps of this procedure are as follows.

Step 1: Minimizing (6.6) over $\mathbf{P}_k$ subject to $\mathbf{P}_k^T\mathbf{P}_k = \mathbf{I}_R$ is equivalent to maximizing

$$\operatorname{tr}\left(\mathbf{F}\mathbf{C}_k\mathbf{B}^T\mathbf{X}_k^T\mathbf{P}_k\right), \tag{6.7}$$

over $\mathbf{P}_k$. The solution of (6.7) is found by computing the singular value decomposition (SVD) of $\mathbf{F}\mathbf{C}_k\mathbf{B}^T\mathbf{X}_k^T$. That is, if $\mathbf{F}\mathbf{C}_k\mathbf{B}^T\mathbf{X}_k^T = \mathbf{U}_k\mathbf{\Delta}_k\mathbf{V}_k^T$ is the SVD, then

$$\mathbf{P}_k = \mathbf{V}_k\mathbf{U}_k^T, \tag{6.8}$$

is the optimal solution. This step is repeated for $k = 1, \ldots, K$.

Step 2: Since $\mathbf{P}_k$, $k = 1, \ldots, K$, are columnwise orthonormal, the problem of minimizing (6.6) over $\mathbf{F}, \mathbf{C}_1, \ldots, \mathbf{C}_K$ and $\mathbf{B}$ reduces to minimizing

$$\sum_{k=1}^{K} \|\mathbf{P}_k^T \mathbf{X}_k - \mathbf{F}\mathbf{C}_k\mathbf{B}^T\|^2. \tag{6.9}$$

Minimizing (6.9) is equivalent to fitting the Parafac model to an $R \times J \times K$ array with frontal slices $\mathbf{P}_k^T \mathbf{X}_k$, $k = 1, \ldots, K$. Therefore, the Parafac ALS algorithm can be applied to find a solution of (6.9).

The procedure with steps 1 and 2 guarantees monotonic convergence of criterion (6.6).

Matrix $\mathbf{B}$ is initialized as the loadings obtained from a PCA of the matrix containing $\mathbf{X}_1, \ldots, \mathbf{X}_K$ stacked below each other, the matrices $\mathbf{F}$ and $\mathbf{C}_1, \ldots, \mathbf{C}_K$ are initialized as identity matrices. The complete direct fitting Parafac2 algorithm can be summarized as follows.

1. Initialize $\mathbf{B}$ as the loading matrix from PCA on $\sum_{k=1}^{K} \mathbf{X}_k^T \mathbf{X}_k$ and initialize $\mathbf{F}$ and $\mathbf{C}_1, \ldots, \mathbf{C}_K$ as $\mathbf{I}_R$.

2. Compute the SVD $\mathbf{F}\mathbf{C}_k\mathbf{B}^T\mathbf{X}_k^T = \mathbf{U}_k\mathbf{\Delta}_k\mathbf{V}_k^T$ and update $\mathbf{P}_k$ as $\mathbf{V}_k\mathbf{U}_k^T$, $k = 1, \ldots, K$.

3. Update $\mathbf{F}$, $\mathbf{B}$ and $\mathbf{C}_1, \ldots, \mathbf{C}_K$ by one run of the Parafac ALS algorithm applied to the $R \times J \times K$ three-way array with frontal slices $\mathbf{P}_k^T \mathbf{X}_k$, $k = 1, \ldots, K$.

4. Evaluate the function value $\sigma_2^{new} = \sum_{k=1}^{K} \|\mathbf{X}_k - \mathbf{P}_k\mathbf{F}\mathbf{C}_k\mathbf{B}^T\|^2$. If $\sigma_2^{old} - \sigma_2^{new} > \varepsilon\sigma_2^{old}$ for some small value ε, repeat 2.

Due to sufficiency of cross-products, there is a optional step to make the algorithm faster in the case $N_k > J$. In this case, $\mathbf{X}_k$ must be replaced by a

$(m_k \times J)$ matrix $\mathbf{H}_k$ with $m_k < N_k$ such that $\mathbf{X}_k^T\mathbf{X}_k = \mathbf{H}_k^T\mathbf{H}_k$ (the Cholesky decomposition of $\mathbf{X}_k^T\mathbf{X}_k$ was used to obtain $\mathbf{H}_k$). For details, see Kiers et al. (1999).

Note that Kiers et al. (1999) use only one iteration of the Parafac ALS algorithm in point 3 above, while we run the Parafac ALS algorithm to convergence. In our simulation study in section 6.4, this yields considerably better recovery results.

6.2.2 Algorithm for indirect Parafac2 via direct Parafac2

To fit the indirect Parafac2 model (6.2), Kiers (1993) provided an ALS algorithm. The following function has to be minimized:

$$\sum_{k=1}^{K} \|\mathbf{\Sigma}_k - \mathbf{B}\mathbf{C}_k\mathbf{\Phi}\mathbf{C}_k\mathbf{B}^T\|^2. \tag{6.10}$$

As mentioned in Kiers et al. (1999), the ALS algorithm of indirect Parafac2 that is proposed in Kiers (1993) converges very slowly. Also, it is rather complicated. Here, we propose an algorithm for fitting indirect Parafac2 that can overcome the limitations of the algorithm in Kiers (1993). Our algorithm is as follows.

Step 1. Compute the SVD or eigendecomposition (EVD) of $\mathbf{\Sigma}_k$ as $\tilde{\mathbf{V}}_k\tilde{\mathbf{C}}_k\tilde{\mathbf{V}}_k^T$. Obtain matrix $\tilde{\mathbf{Q}}_k$ as the first R rows of $\tilde{\mathbf{C}}_k^{\frac{1}{2}}\tilde{\mathbf{V}}_k^T$, corresponding to the R largest singular values.

Step 2. Fit the direct Parafac2 model $\tilde{\mathbf{Q}}_k \approx \mathbf{P}_k\mathbf{F}\mathbf{C}_k\mathbf{B}^T$, $k = 1, \ldots, K$, by means of the direct Parafac2 algorithm of Kiers et al. (1999) (section 6.2.1). Matrix $\mathbf{\Phi}$ is defined as $\mathbf{F}^T\mathbf{F}$.

The fit percentage of the indirect Parafac2 model can be computed as

$$100 - 100 \cdot \frac{\sum_{k=1}^{K} ||\mathbf{\Sigma}_k - \mathbf{B}\mathbf{C}_k\mathbf{\Phi}\mathbf{C}_k\mathbf{B}^T||^2}{\sum_{k=1}^{K} ||\mathbf{\Sigma}_k||^2}. \tag{6.11}$$

In step 2, we run the ALS algorithm of direct Parafac2 10 times for random starting values and 1 time for the starting values suggested by Kiers et al. (1999) in section 6.2.1. We then keep the indirect Parafac2 solution with the highest fit percentage. In our indirect Parafac2 algorithm, we can choose orthogonal or oblique factors because this can be done in Parafac ALS (step 2 in section 6.2.1). The convergence criterion of our indirect Parafac2 algorithm is based on criterion (6.10).

With two relatively simple steps, our procedure is easier to use than the indirect Parafac2 algorithm of Kiers (1993). Moreover, our procedure is also useful in the case one wants to have orthogonal factors. This is not possible in the procedure of Kiers (1993).

Note that $\tilde{\mathbf{Q}}_k^T\tilde{\mathbf{Q}}_k$ with $\tilde{\mathbf{Q}}_k$ as in step 1 above is a best rank-R approximation of $\mathbf{\Sigma}_k$ (Eckart & Young, 1936). As an alternative, in step 1 and step 2 above, we can use $\tilde{\mathbf{Y}}_k = \tilde{\mathbf{C}}_k^{\frac{1}{2}}\tilde{\mathbf{V}}_k^T$ instead of $\tilde{\mathbf{Q}}_k$. Then $\tilde{\mathbf{Y}}_k^T\tilde{\mathbf{Y}}_k = \mathbf{\Sigma}_k$. In the next section, we conduct a simulation study with both these variants of our algorithm for indirect Parafac2.

6.2.3 Simulations for perfect fit

We do partly the same simulations as Kiers (1993) to assess the performance of our indirect Parafac2 algorithm. We generate 160 data sets $\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_K$ as $\mathbf{\Sigma}_k = \mathbf{B}\mathbf{C}_k\mathbf{\Phi}\mathbf{C}_k\mathbf{B}^T$, where matrices $\mathbf{B}, \mathbf{C}_1, \ldots, \mathbf{C}_K$ have random entries from the standard normal distribution, and $\mathbf{\Phi}$ is a positive definite matrix and computed

as the cross-product of a $R \times R$ matrix $\mathbf{F}$ having random entries from the standard normal distribution and column sum of squares equal to 1. Hence, for each data set perfect fit can be attained. We consider $J = 6$ variables for the first 80 sets, and $J = 10$ for the next 80 sets. In both cases we set $K = 3$ for the first 40 sets, and $K = 6$ for the next 40 sets. Half of the 160 data sets feature $R = 2$ factors, and half have $R = 3$. Hence, for each triplet (J, K, R) we generate 20 datasets. We fit our algorithm in section 6.2.2 to each dataset, where we set the convergence criterion to $1e - 7$, and we abort the direct Parafac2 algorithm in step 2 after 100 iterations. We denote $\mathbf{C}$ as a $K \times R$ matrix whose k-th row contains the diagonal of $\mathbf{C}_k$, $k = 1, \ldots, K$. To compare estimated and true values of $\mathbf{B}$ and $\mathbf{C}$, we use congruence coefficients of the true value and estimate of each column of $\mathbf{B}$ and $\mathbf{C}$. For two vectors $\mathbf{h}_1$ and $\mathbf{h}_2$, the congruence coefficient is given by

$$\frac{\mathbf{h}_1^T \mathbf{h}_2}{\sqrt{\mathbf{h}_1^T \mathbf{h}_1}\sqrt{\mathbf{h}_2^T \mathbf{h}_2}}.$$

We obtain results that are better in terms of fit than those in Kiers (1993). Our results are described in Table 6.1.

Each cell in Table 6.1 reports the number of cases (out of 20) with a fit percentage less than 99 % (and less than 99.9% parenthesized). Our results and the results of Kiers (1993) for the case of oblique factors (i.e., random positive definite $\mathbf{\Phi}$) are reported in the fourth column and the last column, respectively. In the results of Kiers (1993), there are 45 data sets (out of 160) with fit percentage less than 99.9% (and one of these has fit percentage less than 99%), whereas all data sets in our simulations have fit percentages larger than or equal to 99.9%. This shows that our estimation procedure performs better than that of Kiers (1993) on the simulated datasets. The same counts are obtained for both variants of the algorithm (i.e., for both $\tilde{\mathbf{Q}}_k$ and $\tilde{\mathbf{Y}}_k$ in step 1 and step 2 of section 6.2.2).

Table 6.1: Fit percentage counts for simulated indirect Parafac2 data and the algorithm of section 2.2. Each cell contains 20 datasets and the number of datasets for which the fit less than 99% (99.9%)

J	K	R	fit % counts	fit % counts for $\mathbf{\Phi} = \mathbf{I}_R$	fit % counts from Kiers (1993)
6	3	2	0(0)	0(0)	0(0)
6	6	2	0(0)	0(0)	0(3)
6	3	3	0(0)	0(0)	0(5)
6	6	3	0(0)	0(0)	1(16)
10	3	2	0(0)	0(0)	0(0)
10	6	2	0(0)	0(0)	0(1)
10	3	3	0(0)	0(0)	0(3)
10	6	3	0(0)	0(0)	0(17)

Due to the uniqueness condition of Ten Berge and Kiers (1996), the Parafac2 solutions with $(J, K, R) \in \{(6, 6, 2), (10, 6, 2)\}$ for the case of oblique factors are unique. For these, we also compute the congruence coefficients between each column of the true $\mathbf{B}$,$\mathbf{C}$ and their estimates, and the mean absolute deviation (MAD) and the standard deviation (SD) for the true factor correlation in $\mathbf{\Phi}$ and its estimates. All congruence coefficients for $(J, K, R) \in \{(6, 6, 2), (10, 6, 2)\}$ are at least 0.95 for both variants of the algorithm. The MAD of the factor correlation estimates for $(J, K, R) = (6, 6, 2)$ is rounded equal to 0.00 with SD = 0.00 if we use $\tilde{\mathbf{Q}}_k$ in section 6.2.2 (and MAD = 0.01 with SD = 0.02 if we use $\tilde{\mathbf{Y}}_k$ in section 6.2.2). The MAD of the factor correlation estimates for $(J, K, R) = (10, 6, 2)$ is 0.00 (rounded) with SD = 0.01 if we use $\tilde{\mathbf{Q}}_k$ in section 6.2.2 (and MAD = 0.01 with SD = 0.01 if we use $\tilde{\mathbf{Y}}_k$ in section 6.2.2).

We did the same simulations for the case of orthogonal factors ($\mathbf{\Phi} = \mathbf{I}_R$). Then we have an Indscal model (Carroll & Chang, 1970), which has a unique solution due to Kruskal's condition (1977)

$$2k_{\mathbf{B}} + k_{\mathbf{C}} \geqslant 2R + 2,$$

where $k_{\mathbf{Z}}$ denotes the k-rank of a matrix $\mathbf{Z}$. The latter is defined as the largest number x such that every subset of x column of $\mathbf{Z}$ is linearly independent. The results are reported in the fifth column of Table 6.1. There is no data set with fit percentage less than 99.9%. Moreover, congruence coefficients between the absolute value of the true $\mathbf{C}$ and the absolute values of its estimates (due to (6.3)) are at least 0.95 for all cases of (J, K, R). The same holds for the estimates of $\mathbf{B}$. This is true for both variants of the algorithm. Hence, the recovery of factor loadings and factor strengths for orthogonal factors is very good.

6.2.4 Simulations for imperfect fit

Here, we assess the performance of the estimation procedure in section 6.2.2 for noisy data. We set $J = 6$, $K = 5$, and $R = 2$. For given $\mathbf{B}, \mathbf{C}, \mathbf{\Phi}$, we generate random data $\mathbf{X}^{(k)}_{(N\times J)}$ with population correlation matrix $\mathbf{\Sigma}^{(k)} = \mathbf{B}\mathbf{C}_k\mathbf{\Phi}\mathbf{C}_k\mathbf{B}^T$. Next, we apply the indirect Parafac2 estimation procedure to the sample covariance matrix of $\mathbf{X}^{(k)}_{(N\times J)}$ for each k. The true matrices $\mathbf{B}, \mathbf{C}$, and $\mathbf{\Phi}$ are

$$\mathbf{B}_1 = \begin{pmatrix} 0.80 & 0.10 \\ 0.10 & 0.83 \\ 0.83 & 0.10 \\ 0.10 & 0.79 \\ 0.83 & 0.10 \\ 0.10 & 0.82 \end{pmatrix}, \qquad \mathbf{B}_2 = \begin{pmatrix} 0.65 & 0.10 \\ 0.10 & 0.63 \\ 0.61 & 0.10 \\ 0.10 & 0.70 \\ 0.64 & 0.10 \\ 0.10 & 0.62 \end{pmatrix}, \qquad \mathbf{C}_1 = \begin{pmatrix} 1.00 & 0.80 \\ 0.80 & 1.20 \\ 1.19 & 0.81 \\ 0.81 & 1.18 \\ 1.20 & 0.79 \end{pmatrix},$$

$$\mathbf{C}_2 = \begin{pmatrix} 1.00 & 0.50 \\ 0.50 & 1.20 \\ 1.19 & 0.51 \\ 0.51 & 1.18 \\ 1.20 & 0.59 \end{pmatrix}, \qquad \mathbf{\Phi}_1 = \begin{pmatrix} 1 & -0.40 \\ -0.40 & 1 \end{pmatrix}.$$

Hence, for $\mathbf{C}$ we use true matrices $\mathbf{C}_1$ and $\mathbf{C}_2$, and for $\mathbf{B}$ we use true matrices $\mathbf{B}_1$ and $\mathbf{B}_2$. We consider both orthogonal factors ($\mathbf{\Phi} = \mathbf{I}_2$) and oblique factors ($\mathbf{\Phi} = \mathbf{\Phi}_1$). Note that the Parafac2 uniqueness condition of Ten Berge and Kiers (1996) is satisfied for the true Parafac2 models. Hence, nonuniqueness cannot hamper the recovery of the true loadings by our algorithm.

For the sample size N, we take $N = 100$. The data are generated as

$$\mathbf{X}^{(k)}_{(N\times J)} = \mathbf{Z}^{(k)}_{(N\times J)} \left(\mathbf{\Sigma}^{(k)}\right)^{\frac{1}{2}}, \quad k = 1 \ldots K, \tag{6.12}$$

where $\mathbf{Z}^{(k)}_{(N\times J)}$ has random entries from the standard normal distribution, and $\mathbf{\Sigma}^{(k)}$ is the population correlation matrix. For each choice of true model, we generate 100 datasets as in (6.12) and fit the indirect Parafac2 model to the sample covariance matrix. In the direct Parafac2 algorithm in step 2, we use convergence criterion $1e-7$ and 100 iterations. We compare the true values of $\mathbf{B}$ and $\mathbf{C}$ to their estimates by means of congruence coefficients for each column of $\mathbf{B}$ and $\mathbf{C}$. Also, we report the mean absolute deviation (MAD) and mean bias (BIAS) for the estimates of $\mathbf{B}$ and $\mathbf{C}$. For estimate $\hat{\mathbf{B}}$, the MAD and BIAS are defined as

$$\text{MAD}(\hat{\mathbf{B}}, \mathbf{B}^{(\text{true})}) = \frac{\sum_{j=1}^{J}\sum_{r=1}^{R} |\hat{b}_{jr} - b^{(\text{true})}_{jr}|}{JR},$$

$$\text{BIAS}(\hat{\mathbf{B}}, \mathbf{B}^{(\text{true})}) = \frac{\sum_{j=1}^{J}\sum_{r=1}^{R} (\hat{b}_{jr} - b^{(\text{true})}_{jr})}{JR},$$

where the columns of $\hat{\mathbf{B}}$ are rescaled such that the column sum-of-squares are the same as for the true $\mathbf{B}$. The MAD and BIAS are computed analogously for estimates of $\mathbf{C}$.

For each true model, we use both orthogonal factors and oblique factors in our algorithm. When using orthogonal estimation, we take the absolute value of the estimate of $\mathbf{C}$ due to (6.3). When using oblique factors in the estimation, the factor correlations are often small. Also in these cases we take the absolute value of the estimate of $\mathbf{C}$. This yields larger congruence coefficients between the true $\mathbf{C}$ and its estimate. Apparently, the estimated oblique factors are close enough to being orthogonal to result in uniqueness up to sign for $\mathbf{C}$ as in (6.3).

Table 6.2 contains the mean and standard deviation of the congruence coefficients of the columns of $\mathbf{B}$ and $\mathbf{C}$ for each case, and the mean and standard deviation of the MAD and BIAS values for each case. We only report the results when using $\tilde{\mathbf{Q}}_k$ in step 1 of the algorithm. The results for $\tilde{\mathbf{Y}}_k$ are very similar. In Table 6.2 it can be seen that the recovery of $\mathbf{B}$ is very good in general, and the recovery of $\mathbf{C}$ is excellent. The recovery of $\mathbf{B}$ is worse for oblique estimation. The estimates of $\mathbf{C}$ are not biased, while the estimates of $\mathbf{B}$ are slightly negatively biased.

When oblique factors are used in the estimation, the number of cases with diverging components in the Candecomp/Parafac algorithm is also reported in Table 6.2. We define two components as diverging if their congruence coefficient is smaller than -0.90. If diverging components occur, then a best-fitting Candecomp/Parafac model probably does not exist (Krijnen et al., 2008; Stegeman, 2008, 2012). As can be seen, diverging components occur most when $\mathbf{B}_2$ and $\mathbf{C}_2$ are the true matrices, which contain the smallest coefficients. In this case there is more noise in the generated data, and it is known that random data may often yield diverging components when CP is fitted; see Stegeman (2006, 2008, 2012).

Table 6.2: Mean and standard deviation of congruence coefficients (CC), MAD and BIAS between estimate and true values of $\mathbf{B}$, $\mathbf{C}$, for noisy data and using $\tilde{\mathbf{Q}}_k$ in the indirect Parafac2 algorithm, and sample size $N = 100$.

$\mathbf{C}$	$\mathbf{B}$	$\mathbf{\Phi}$	estimation	$\mathbf{C}$				$\mathbf{B}$				div.
				CC		MAD	BIAS	CC		MAD	BIAS	comp.
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	orthogonal	1.00	1.00	0.05	-0.01	0.99	1.00	0.03	-0.00	-
				(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	oblique	1.00	1.00	0.05	-0.00	0.97	0.98	0.09	-0.01	0
				(0.00)	(0.00)	(0.01)	(0.01)	(0.04)	(0.03)	(0.06)	(0.08)	
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	orthogonal	1.00	1.00	0.04	-0.00	1.00	1.00	0.01	-0.00	-
				(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	oblique	1.00	1.00	0.06	0.01	0.93	0.94	0.12	-0.02	0
				(0.00)	(0.00)	(0.01)	(0.00)	(0.05)	(0.04)	(0.04)	(0.10)	
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	orthogonal	1.00	1.00	0.05	-0.00	0.98	0.98	0.08	-0.07	-
				(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	oblique	1.00	1.00	0.05	0.00	0.94	0.93	0.14	-0.05	1
				(0.00)	(0.00)	(0.01)	(0.01)	(0.06)	(0.06)	(0.07)	(0.14)	
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	orthogonal	1.00	1.00	0.05	-0.01	0.98	0.98	0.06	-0.05	-
				(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	oblique	1.00	1.00	0.05	0.00	0.92	0.93	0.12	-0.05	2
				(0.00)	(0.00)	(0.01)	(0.01)	(0.07)	(0.06)	(0.06)	(0.11)	
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{I}_2$	orthogonal	1.00	1.00	0.04	-0.00	1.00	1.00	0.01	-0.00	-
				(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{I}_2$	oblique	0.99	0.99	0.09	0.02	0.93	0.94	0.14	-0.02	0
				(0.00)	(0.01)	(0.04)	(0.01)	(0.05)	(0.04)	(0.08)	(0.13)	
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{I}_2$	orthogonal	1.00	1.00	0.04	-0.00	1.00	1.00	0.01	-0.00	-
				(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{I}_2$	oblique	0.99	0.99	0.10	0.02	0.93	0.94	0.12	-0.02	0
				(0.00)	(0.01)	(0.03)	(0.01)	(0.05)	(0.04)	(0.06)	(0.10)	
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	orthogonal	1.00	1.00	0.04	-0.01	0.99	0.99	0.06	-0.06	-
				(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	oblique	0.99	0.99	0.09	0.02	0.94	0.95	0.14	-0.02	5
				(0.01)	(0.01)	(0.04)	(0.02)	(0.06)	(0.05)	(0.08)	(0.12)	
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	orthogonal	1.00	1.00	0.04	-0.01	0.99	0.99	0.05	-0.04	-
				(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	oblique	0.99	0.99	0.08	0.01	0.92	0.94	0.11	-0.05	11
				(0.01)	(0.01)	(0.04)	(0.02)	(0.07)	(0.05)	(0.06)	(0.11)	

6.3 Multi-set factor analysis by means of Parafac2

Here, we present our algorithm to estimate the multi-set Parafac2 factor model (6.2). Recall from section 6.1 that we first use Minimum Rank Factor Analysis (MRFA) to estimate the unique variances $\mathbf{U}_k$ for each k. Next, we apply our indirect Parafac2 algorithm in section 6.2.2 to $\mathbf{\Sigma}_k - \mathbf{U}_k$, $k = 1, \ldots, K$, to obtain estimates for $\mathbf{B}$, $\mathbf{\Phi}$, and $\mathbf{C}_1, \ldots, \mathbf{C}_K$. In section 6.3.1, we briefly describe MRFA. In section 6.3.2, we present and discuss our algorithm for the multi-set Parafac2 factor model.

6.3.1 Minimum Rank Factor Analysis

Here, we briefly describe the MRFA method for two-mode factor analysis. The covariance model is here give by $\mathbf{\Sigma} \approx \mathbf{B\Phi B}^T + \mathbf{U}$, where $\mathbf{\Sigma}$ is the $J \times J$ data covariance matrix, $\mathbf{B}$ is the $J \times R$ loading matrix, and $\mathbf{U}$ is the $J \times J$ diagonal matrix of unique variances. MRFA is used to estimate $(\mathbf{B}, \mathbf{\Phi}, \mathbf{U})$; see Ten Berge and Kiers (1991). The MRFA algorithm computes the unique variances $\mathbf{U}$ such that $\mathbf{U}$ is nonnegative, $\mathbf{\Sigma} - \mathbf{U}$ is a covariance matrix, and the unexplained common variance in $\mathbf{\Sigma} - \mathbf{U} \approx \mathbf{B\Phi B}^T$ is minimized. Since $\mathbf{\Sigma} - \mathbf{U}$ is a covariance matrix, all its eigenvalues are nonnegative. Then the best approximation $\mathbf{B\Phi B}^T$ is obtained from the R largest eigenvalues and associated eigenvectors of $\mathbf{\Sigma} - \mathbf{U}$, and the minimum unexplained common variance in $\mathbf{\Sigma} - \mathbf{U} \approx \mathbf{B\Phi B}^T$ is equal to the sum of the $J - R$ smallest eigenvalues of $\mathbf{\Sigma} - \mathbf{U}$; see Eckart and Young (1936).

6.3.2 Estimation procedure for the multi-set Parafac2 factor model

We present the following procedure to estimate our multi-set Parafac2 factor model (6.2). We need to minimize

$$\sum_{k=1}^{K} ||\mathbf{\Sigma}_k - \mathbf{U}_k - \mathbf{B}\mathbf{C}_k\mathbf{\Phi}\mathbf{C}_k\mathbf{B}^T||^2. \tag{6.13}$$

The steps of our estimation procedure are as follows.

Step 1. For each $k \in \{1, .., K\}$, use the MRFA method of Ten Berge and Kiers (1991) to estimate the unique variances $\mathbf{U}_k$ corresponding to $\mathbf{\Sigma}_k$.

Step 2. Compute the EVD or SVD $(\mathbf{\Sigma}_k - \mathbf{U}_k) = \mathbf{V}_k\mathbf{C}_k\mathbf{V}_k^T$, with $\mathbf{V}_k$ having orthonormal columns, and $\mathbf{C}_k$ the diagonal matrix containing the eigenvalues in decreasing order. Set $\mathbf{Y}_k = \mathbf{C}_k^{\frac{1}{2}}\mathbf{V}_k^T$. Hence, $(\mathbf{\Sigma}_k - \mathbf{U}_k) = \mathbf{Y}_k^T\mathbf{Y}_k$, $k = 1, \ldots K$. If $\mathbf{Q}_k$ contains the first R rows of $\mathbf{C}_k^{\frac{1}{2}}\mathbf{V}_k^T$, then $\mathbf{Q}_k^T\mathbf{Q}_k$ is a best rank-R approximation of $\mathbf{\Sigma}_k - \mathbf{U}_k$.

Step 3. Fit the direct Parafac2 model $\mathbf{Y}_k \approx \mathbf{P}_k\mathbf{F}\mathbf{C}_k\mathbf{B}^T$, $k = 1, \ldots, K$, by means of the direct Parafac2 algorithm of Kiers et al. (1999) (section 6.2.1). Matrix $\mathbf{\Phi}$ is defined as $\mathbf{F}^T\mathbf{F}$.

Due to the MRFA method of Ten Berge and Kiers (1991) all unique variances $\mathbf{U}_k$ are nonnegative, and $\mathbf{\Sigma}_k - \mathbf{U}_k$ are covariance matrices (i.e., positive semi-definite). By using the ALS algorithm of Kiers et al. (1999) (in step 3), we obtain $\mathbf{Y}_k \approx \mathbf{P}_k\mathbf{F}\mathbf{C}_k\mathbf{B}^T$. This implies that

$$(\mathbf{\Sigma}_k - \mathbf{U}_k) \approx \mathbf{B}\mathbf{C}_k\mathbf{F}^T\mathbf{P}_k^T\mathbf{P}_k\mathbf{F}\mathbf{C}_k\mathbf{B}^T = \mathbf{B}\mathbf{C}_k\mathbf{F}^T\mathbf{F}\mathbf{C}_k\mathbf{B}^T.$$

The percentage of explained common variance for each sample k can be computed as

$$100 - 100 \cdot \frac{\text{ssq}(\mathbf{Y}_k - \mathbf{P}_k\mathbf{F}\mathbf{C}_k\mathbf{B}^T)}{\text{ssq}(\mathbf{Y}_k)}, \quad k = 1 \ldots K, \tag{6.14}$$

where ssq($\mathbf{Z}$) denotes the sum of squares of matrix $\mathbf{Z}$. In the ALS algorithm of step 3 above, $\mathbf{P}_k\mathbf{F}\mathbf{C}_k\mathbf{B}^T$ is the regression of $\mathbf{Y}_k$ on $\mathbf{P}_k\mathbf{F}\mathbf{C}_k$. Since the regression and the residual are orthogonal, it follows that (6.14) can be written as

$$100 \cdot \frac{\text{ssq}(\mathbf{P}_k\mathbf{F}\mathbf{C}_k\mathbf{B}^T)}{\text{ssq}(\mathbf{Y}_k)} = 100 \cdot \frac{\text{trace}(\mathbf{B}\mathbf{C}_k\mathbf{\Phi}\mathbf{C}_k\mathbf{B}^T)}{\text{trace}(\mathbf{\Sigma}_k - \mathbf{U}_k)}, \quad k = 1 \ldots K. \tag{6.15}$$

If the factors are chosen orthogonal, then we may also obtain a percentage of explained common variance due to each factor for each sample k. Namely, we then have $\mathbf{\Phi} = \mathbf{I}_R$ in (6.15), and we can define the explained common variance due to factor r for sample k as the sum of squares of column r of matrix $\mathbf{B}\mathbf{C}_k$. If we denote $\mathbf{b}_r$ as column r of matrix $\mathbf{B}$ and $c_{k,r}$ as the r-th diagonal element of $\mathbf{C}_k$, then column r of $\mathbf{B}\mathbf{C}_k$ is $c_{k,r}\mathbf{b}_r$ and has sum of squares equal to $c_{k,r}^2\mathbf{b}_r^T\mathbf{b}_r$. Hence, we have

$$100 \cdot \frac{\text{trace}(\mathbf{B}\mathbf{C}_k^2\mathbf{B}^T)}{\text{trace}(\mathbf{\Sigma}_k - \mathbf{U}_k)} = \sum_{r=1}^{R} \left(100 \cdot \frac{c_{k,r}^2\mathbf{b}_r^T\mathbf{b}_r}{\text{trace}(\mathbf{\Sigma}_k - \mathbf{U}_k)} \right), \quad k = 1 \ldots K, \tag{6.16}$$

where the summands in the right-hand side express the percentage of explained common variance due to each factor for each sample k.

Denote $\mathbf{w}_{k,m}$ as column m of $\mathbf{Y}_k - \mathbf{P}_k\mathbf{F}\mathbf{C}_k\mathbf{B}^T$, where $\mathbf{Y}_k$ is obtained from step 2 as above, and $\mathbf{B}$, $\mathbf{C}_1, \ldots, \mathbf{C}_K$, $\mathbf{F}$, and $\mathbf{P}_1, \ldots, \mathbf{P}_K$ are obtained from step 3 as above. We define the percentage of explained common variance for each variable j and each sample k as

$$100 - 100 \cdot \frac{\text{ssq}\left(\mathbf{w}_{k,m}\right)}{\text{ssq}\left(\mathbf{\Sigma}_k - \mathbf{U}_k\right)_{jj}}. \tag{6.17}$$

If we use $\mathbf{Q}_k$ in step 2 instead of $\mathbf{Y}_k$, then we obtain an alternative algorithm where we can only use (6.14) and not (6.15) to compute the percentage of explained common variance for each sample k. And thus we cannot use (6.17) to compute the percentage of explained common variance for each variable and each sample.

Note that our estimation procedure does not minimize (6.13) completely, since the estimation of $\mathbf{U}_k$ by MRFA does not take into account the Parafac2 model form for $\mathbf{\Sigma}_k - \mathbf{U}_k$. Instead, MRFA assumes a separate R-factor model for each sample k. It would be consistent with this assumption to use $\mathbf{Q}_k$ in step 3, since $\mathbf{Q}_k^T\mathbf{Q}_k$ is a best rank-R approximation of $\mathbf{\Sigma}_k - \mathbf{U}_k$. However, as explained above, if we want to compute explained common variances as in (6.17), we need to use $\mathbf{Y}_k$ in step 3. As we will see in the simulation study in the next section, the recovery results when using $\mathbf{Q}_k$ or $\mathbf{Y}_k$ are very similar.

Also, the fit of direct Parafac2 to $\mathbf{Q}_1, \ldots, \mathbf{Q}_K$ in step 3 is very high in our simulations in section 6.4 (above 98 percent) and in our application in section 6.5 (above 93 percent). This reduced the need for a simultaneous estimation method for the unique variances and the Parafac2 matrices.

6.4 Simulation study

In this section, we assess the performance of the estimation procedure in section 6.3.2. We use the same setup as in section 6.2.4, and use the same true $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{\Phi}$ matrices. For given $\mathbf{B}, \mathbf{C}, \mathbf{\Phi}$, and $\mathbf{U}_k$ for $k = 1 \ldots K$, we generate random data $\mathbf{X}^{(k)}_{(N\times J)}$ with population correlation matrix $\mathbf{\Sigma}^{(k)}$ satisfying (6.2). Next, we apply the estimation procedure of section 6.3.2 to the sample covariance matrix of $\mathbf{X}^{(k)}_{(N\times J)}$ for each k. After the true $\mathbf{B}, \mathbf{C}, \mathbf{\Phi}$ are chosen, the unique variances on the diagonal of $\mathbf{U}_k$ are determined such that the population correlation matrices $\mathbf{\Sigma}^{(k)}$ in (6.2) have ones on their diagonals. For the sample size N, we take $N = 100$. The data are generated as in (6.12). For each choice of true model, we generate 100 datasets and fit the multi-set Parafac2 factor model (6.2) to the sample covariance matrix. In our algorithm in section 6.3.2, we use convergence criterion $1e-7$ and 100 iterations for the direct Parafac2 algorithm in step 3.

As in section 6.2.4, we take the absolute value of the estimates of $\mathbf{C}$.

We compare the true values of $\mathbf{B}$ and $\mathbf{C}$ to their estimates by means of congruence coefficients for each column of $\mathbf{B}$ and $\mathbf{C}$. Also, we report the mean absolute deviation (MAD) and mean bias (BIAS) for the estimates of $\mathbf{B}$ and $\mathbf{C}$. Table 6.3 reports the mean and standard deviations of the congruence coefficients when matrix $\mathbf{Q}_k$ in step 2 in section 6.3.2 is used (with $\mathbf{Q}_k^T\mathbf{Q}_k$ being a best rank-R approximation of $\mathbf{\Sigma}_k - \mathbf{U}_k$). Table 6.4 reports the mean and standard deviations of the congruence coefficients results when matrix $\mathbf{Y}_k$ in step 2 in section 6.3.2 is used (with $\mathbf{\Sigma}_k - \mathbf{U}_k = \mathbf{Y}_k^T\mathbf{Y}_k$).

In Table 6.5, we report the mean and standard deviation of the mean absolute deviation (MAD) and mean bias (BIAS) of the estimated $\mathbf{B}$ and $\mathbf{C}$ matrices. Only the results when using $\mathbf{Q}_k$ in step 2 in section 6.3.2 are reported. The results when using $\mathbf{Y}_k$ are very similar, yet only slightly worse in oblique estimation cases. Table 6.6 contains the MAD and BIAS values for the estimated unique variance matrices $\mathbf{U}_k$.

The recovery of the true loadings is very good in general, and somewhat better for $\mathbf{C}$ than for $\mathbf{B}$. The recovery is better when the true model has orthogonal factors compared to when the true model has oblique factors, especially for matrix $\mathbf{B}$. Also, using orthogonal factors in the estimation improves the recovery when the true model has orthogonal factors. The estimates of $\mathbf{C}$ are not biased, while the estimates of $\mathbf{B}$ are slightly negatively biased. This is similar to the simulation results for the three-mode factor model in section 4.4. Also similar is the negative bias when estimating the unique variances.

When oblique factors are used in the estimation, the number of cases with diverging components in the Candecomp/Parafac algorithm is also reported. As can be seen in Tables 6.3 and 6.4, there are only very few cases of diverging components.

Table 6.3: Recovery results for sample size $N = 100$ and using $\mathbf{Q}_k$ in the algorithm of section 6.3.2 (with $\mathbf{Q}_k^T\mathbf{Q}_k$ a best rank-R approximation of $\mathbf{\Sigma}_k - \mathbf{U}_k$).

C	**B**	**Φ**	communality	estimation	congr.coeff. **B**		congr.coeff. **C**		div.comp.
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	0.56 0.70	orthogonal	1.00	1.00	1.00	1.00	-
			0.70 0.69 0.70		(0.00)	(0.00)	(0.00)	(0.00)	
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	0.56 0.70	oblique	0.95	0.95	1.00	1.00	0
			0.70 0.69 0.70		(0.07)	(0.06)	(0.00)	(0.00)	
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	0.34 0.44	orthogonal	0.99	0.99	1.00	1.00	-
			0.43 0.44 0.43		(0.01)	(0.01)	(0.00)	(0.00)	
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	0.34 0.44	oblique	0.96	0.96	1.00	1.00	0
			0.43 0.44 0.43		(0.06)	(0.06)	(0.01)	(0.01)	
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	0.50 0.64	orthogonal	0.98	0.98	1.00	1.00	-
			0.64 0.63 0.64		(0.02)	(0.02)	(0.00)	(0.00)	
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	0.50 0.64	oblique	0.93	0.94	1.00	1.00	1
			0.64 0.63 0.64		(0.08)	(0.07)	(0.00)	(0.00)	
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	0.30 0.40	orthogonal	0.97	0.97	1.00	1.00	-
			0.38 0.39 0.38		(0.02)	(0.02)	(0.01)	(0.00)	
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	0.30 0.39	oblique	0.94	0.95	1.00	1.00	1
			0.38 0.39 0.38		(0.08)	(0.06)	(0.01)	(0.01)	
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{I}_2$	0.43 0.57	orthogonal	1.00	1.00	1.00	1.00	-
			0.57 0.56 0.61		(0.00)	(0.00)	(0.00)	(0.00)	
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{I}_2$	0.43 0.57	oblique	0.97	0.97	0.99	0.99	0
			0.57 0.56 0.61		(0.06)	(0.06)	(0.01)	(0.01)	
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{I}_2$	0.26 0.36	orthogonal	0.99	0.99	0.99	0.99	-
			0.35 0.36 0.37		(0.01)	(0.01)	(0.01)	(0.01)	
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{I}_2$	0.26 0.36	oblique	0.97	0.97	0.99	0.98	0
			0.35 0.36 0.37		(0.04)	(0.03)	(0.01)	(0.02)	
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	0.39 0.53	orthogonal	0.98	0.99	1.00	1.00	-
			0.53 0.52 0.56		(0.01)	(0.01)	(0.00)	(0.01)	
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	0.39 0.53	oblique	0.96	0.97	1.00	0.99	0
			0.53 0.52 0.57		(0.05)	(0.05)	(0.00)	(0.01)	
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	0.23 0.33	orthogonal	0.97	0.97	0.99	0.99	-
			0.32 0.32 0.34		(0.02)	(0.02)	(0.01)	(0.01)	
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	0.23 0.33	oblique	0.95	0.95	0.99	0.98	1
			0.32 0.32 0.34		(0.06)	(0.05)	(0.01)	(0.01)	

Table 6.4: Recovery results for sample size $N = 100$ and using $\mathbf{Y}_k$ in the algorithm of section 6.3.2 (with $\mathbf{Y}_k^T\mathbf{Y}_k = \mathbf{\Sigma}_k - \mathbf{U}_k$).

C	**B**	**Φ**	communality	estimation	congr.coeff. **B**		congr.coeff. **C**		div.comp.
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	0.56 0.70	orthogonal	1.00	1.00	1.00	1.00	-
			0.70 0.69 0.70		(0.01)	(0.00)	(0.00)	(0.00)	
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	0.56 0.70	oblique	0.94	0.95	1.00	1.00	0
			0.70 0.69 0.70		(0.07)	(0.06)	(0.00)	(0.00)	
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	0.34 0.44	orthogonal	1.00	1.00	1.00	1.00	-
			0.43 0.44 0.43		(0.01)	(0.01)	(0.00)	(0.00)	
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	0.34 0.44	oblique	0.95	0.96	1.00	1.00	0
			0.43 0.44 0.43		(0.06)	(0.05)	(0.00)	(0.00)	
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	0.50 0.64	orthogonal	1.00	1.00	1.00	1.00	-
			0.64 0.63 0.64		(0.01)	(0.01)	(0.00)	(0.00)	
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	0.50 0.64	oblique	0.94	0.94	1.00	1.00	0
			0.64 0.63 0.64		(0.07)	(0.06)	(0.00)	(0.00)	
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	0.30 0.40	orthogonal	0.97	0.97	1.00	1.00	-
			0.38 0.39 0.38		(0.02)	(0.02)	(0.00)	(0.00)	
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	0.30 0.39	oblique	0.94	0.94	1.00	1.00	1
			0.38 0.39 0.38		(0.07)	(0.07)	(0.00)	(0.00)	
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{I}_2$	0.43 0.57	orthogonal	1.00	1.00	1.00	1.00	-
			0.57 0.56 0.61		(0.00)	(0.00)	(0.00)	(0.00)	
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{I}_2$	0.43 0.57	oblique	0.95	0.96	0.99	0.99	0
			0.57 0.56 0.61		(0.07)	(0.06)	(0.01)	(0.01)	
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{I}_2$	0.26 0.36	orthogonal	0.99	0.99	0.99	0.99	-
			0.35 0.36 0.37		(0.01)	(0.01)	(0.01)	(0.01)	
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{I}_2$	0.26 0.36	oblique	0.96	0.96	0.99	0.99	0
			0.35 0.36 0.37		(0.05)	(0.04)	(0.01)	(0.01)	
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	0.39 0.53	orthogonal	0.99	0.98	1.00	1.00	-
			0.53 0.52 0.56		(0.01)	(0.01)	(0.00)	(0.00)	
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	0.39 0.53	oblique	0.95	0.95	0.99	0.99	1
			0.53 0.52 0.57		(0.06)	(0.06)	(0.01)	(0.01)	
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	0.23 0.33	orthogonal	0.98	0.98	0.99	0.99	-
			0.32 0.32 0.34		(0.01)	(0.01)	(0.01)	(0.01)	
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	0.23 0.33	oblique	0.95	0.95	0.99	0.98	0
			0.32 0.32 0.34		(0.06)	(0.06)	(0.02)	(0.02)	

Table 6.5: Mean and standard deviation of MAD and BIAS between estimated and true values of $\mathbf{B}$ and $\mathbf{C}$, when using $\mathbf{Q}_k$ in the algorithm of section 6.3.2 (with $\mathbf{Q}_k^T\mathbf{Q}_k$ a best rank-R approximation of $\mathbf{\Sigma}_k - \mathbf{U}_k$), and sample size $N = 100$.

$\mathbf{C}$	$\mathbf{B}$	$\mathbf{\Phi}$	estimation	MAD		BIAS	
				$\mathbf{C}$	$\mathbf{B}$	$\mathbf{C}$	$\mathbf{B}$
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	orthogonal	0.02	0.03	0.00	0.00
				(0.00)	(0.01)	(0.00)	(0.00)
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	oblique	0.04	0.12	0.00	-0.02
				(0.08)	(0.02)	(0.10)	(0.00)
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	orthogonal	0.05	0.05	0.00	0.00
				(0.01)	(0.02)	(0.00)	(0.00)
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	oblique	0.06	0.07	0.00	-0.01
				(0.01)	(0.04)	(0.00)	(0.04)
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	orthogonal	0.03	0.09	0.00	-0.07
				(0.01)	(0.01)	(0.00)	(0.01)
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	oblique	0.04	0.13	0.00	-0.09
				(0.02)	(0.08)	(0.00)	(0.10)
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	orthogonal	0.05	0.08	0.00	-0.06
				(0.01)	(0.01)	(0.00)	(0.01)
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	oblique	0.06	0.10	0.00	-0.06
				(0.01)	(0.04)	(0.00)	(0.04)
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{I}_2$	orthogonal	0.04	0.03	0.00	0.00
				(0.01)	(0.01)	(0.00)	(0.00)
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{I}_2$	oblique	0.07	0.09	0.01	-0.01
				(0.05)	(0.07)	(0.01)	(0.07)
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{I}_2$	orthogonal	0.09	0.05	0.01	0.00
				(0.02)	(0.01)	(0.01)	(0.00)
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{I}_2$	oblique	0.11	0.08	0.01	-0.01
				(0.03)	(0.04)	(0.01)	(0.05)
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	orthogonal	0.04	0.08	0.00	-0.06
				(0.01)	(0.01)	(0.01)	(0.01)
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	oblique	0.07	0.12	0.01	-0.08
				(0.04)	(0.07)	(0.01)	(0.09)
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	orthogonal	0.10	0.08	0.00	-0.04
				(0.02)	(0.01)	(0.02)	(0.01)
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	oblique	0.11	0.10	0.01	-0.04
				(0.05)	(0.04)	(0.01)	(0.04)

Table 6.6: Mean and standard deviation of MAD and BIAS between estimated and true values of $\mathbf{U}_k$ for sample size $N = 100$.

$\mathbf{C}$	$\mathbf{B}$	$\mathbf{\Phi}$	estimation	MAD					BIAS				
				$\mathbf{U}_1$	$\mathbf{U}_2$	$\mathbf{U}_3$	$\mathbf{U}_4$	$\mathbf{U}_5$	$\mathbf{U}_1$	$\mathbf{U}_2$	$\mathbf{U}_3$	$\mathbf{U}_4$	$\mathbf{U}_5$
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	orthogonal	0.09	0.06	0.05	0.06	0.05	-0.05	-0.03	-0.02	-0.02	-0.01
				(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{I}_2$	oblique	0.10	0.05	0.05	0.06	0.05	-0.06	-0.02	-0.02	-0.02	-0.02
				(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	orthogonal	0.15	0.12	0.12	0.12	0.13	-0.11	-0.09	-0.08	-0.09	-0.09
				(0.03)	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{I}_2$	oblique	0.14	0.12	0.12	0.12	0.12	-0.09	-0.09	-0.08	-0.08	-0.08
				(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.05)	(0.04)	(0.04)	(0.04)	(0.03)
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	orthogonal	0.11	0.07	0.07	0.07	0.08	-0.07	-0.04	-0.03	-0.04	-0.04
				(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.02)	(0.03)
$\mathbf{C}_1$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	oblique	0.10	0.07	0.07	0.07	0.07	-0.07	-0.03	-0.03	-0.04	-0.04
				(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)	(0.03)	(0.02)	(0.02)
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	orthogonal	0.17	0.13	0.14	0.13	0.13	-0.13	-0.09	-0.10	-0.09	-0.10
				(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)
$\mathbf{C}_1$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	oblique	0.16	0.14	0.14	0.14	0.13	-0.12	-0.10	-0.10	-0.10	-0.10
				(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{I}_2$	orthogonal	0.14	0.08	0.09	0.09	0.07	-0.11	-0.06	-0.06	-0.07	-0.04
				(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)	(0.03)
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{I}_2$	oblique	0.13	0.09	0.09	0.10	0.07	-0.10	-0.06	-0.06	-0.07	-0.03
				(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.05)	(0.04)	(0.03)	(0.04)	(0.03)
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{I}_2$	orthogonal	0.18	0.16	0.16	0.16	0.15	-0.16	-0.14	-0.14	-0.13	-0.12
				(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{I}_2$	oblique	0.19	0.15	0.16	0.16	0.15	-0.17	-0.12	-0.13	-0.14	-0.12
				(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	orthogonal	0.14	0.11	0.10	0.11	0.09	-0.11	-0.09	-0.08	-0.09	-0.06
				(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
$\mathbf{C}_2$	$\mathbf{B}_1$	$\mathbf{\Phi}_1$	oblique	0.13	0.10	0.10	0.10	0.09	-0.11	-0.08	-0.08	-0.08	-0.06
				(0.05)	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	orthogonal	0.18	0.17	0.17	0.18	0.16	-0.16	-0.15	-0.15	-0.15	-0.13
				(0.05)	(0.05)	(0.06)	(0.05)	(0.04)	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)
$\mathbf{C}_2$	$\mathbf{B}_2$	$\mathbf{\Phi}_1$	oblique	0.20	0.17	0.17	0.17	0.16	-0.17	-0.16	-0.14	-0.14	-0.14
				(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)	(0.06)

When using $\mathbf{Q}_k$ in the algorithm, estimation accuracy in terms of MAD or congruence is slightly better for cases of oblique estimation, and estimation accuracy in terms of congruence is slightly worse for cases of orthogonal estimation. Hence, it seems to make very little difference whether $\mathbf{Q}_k$ or $\mathbf{Y}_k$ are used in the algorithm.

6.5 Application of the multi-set Parafac2 factor model

We analyse real multi-set data from Meijer, Egberink, Emons, and Sijtsma (2008) concerning the Self-Perception Profile for Children (SPPC). The SPPC is used to investigate the judgement of children between 8 and 12 years of age about their own functioning in several specific domains and their global self-worth. The SPPC consists of six subscales each consisting of six items. Five of the six subscales represent specific domains of self-concept: Scholastic Competence (SC), Social Acceptance (SA), Athletic Competence (AC), Physical Appearance (PA), and Behavioral Conduct (BC). The sixth scale measures Global Self-worth (GS), which is a more general concept. When a child fills out the SPPC, he or she first chooses which of two statements applied to him or her and then indicates if the chosen statement is "sort of true for me"or "really true for me". Scoring is done on a 4-point scale ranging from 1 to 4, where the answer least indicative of competence is scored 1, and the answer most indicative of competence is scored 4. Meijer et al. (2008) obtained a survey on a sample of $N = 611$ children between 7 and 13 years of age, with 343 girls and 268 boys.

Meijer et al. (2008) showed that there are differences in the IRT model fit for children between age 8 through 9 and for children between age 10 through 13, and between boys and girls. These differences may be due to young children finding the questions too difficult or not yet having a differentiated self-concept.

Also, girls may have a more differentiated self-concept than boys. Therefore, we divide this sample into $K = 4$ groups: young girls (YoGi) consists 147 girls with the age between 8 and 9 years, young boys (YoBo) consists 119 boys with the age between 8 and 9 years, old girls (OlGi) consists 196 girls with the age between 10 and 13 years, and old boys (OlBo) consists 149 boys with the age between 10 and 13 years.

Since analysing all thirty-six items of SPPC does not yield meaningful results, we use the sumscores on the six subscales instead. For this dataset, we have four 6×6 covariance matrices $\mathbf{\Sigma}_1$, $\mathbf{\Sigma}_2$, $\mathbf{\Sigma}_3$, $\mathbf{\Sigma}_4$, which are given in the Appendix. We apply our multi-set Parafac2 factor model (6.2) to this dataset, using the estimation procedure in section 3.2 with $\mathbf{Y}_k$ in step 2. In this dataset, we have $J = 6$, $K = 4$. First, we estimate the multi-set Parafac2 factor model (6.2) with $R = 2$ factors.

The results for $R = 2$ orthogonal factors are as follows

$$\mathbf{B} = \begin{pmatrix} 0.21 & \mathbf{0.48} \\ 0.34 & \mathbf{0.58} \\ 0.21 & \mathbf{0.62} \\ \mathbf{0.58} & -0.03 \\ 0.29 & 0.21 \\ \mathbf{0.61} & 0.02 \end{pmatrix} \begin{matrix} \text{SC} \\ \text{SA} \\ \text{AC} \\ \text{PA} \\ \text{BC} \\ \text{GS} \end{matrix}, \qquad \mathbf{C} = \begin{pmatrix} 5.79 & 3.18 \\ 4.95 & 4.19 \\ 6.05 & 3.12 \\ 4.80 & 2.91 \end{pmatrix} \begin{matrix} \text{YoGi} \\ \text{YoBo} \\ \text{OlGi} \\ \text{OlBo} \end{matrix}, \tag{6.18}$$

where $\mathbf{B}$ has column sum of squares equal to 1, and the loading whose absolute values are larger than or equal to 0.4 are in bold font.

The unique variances for each k are given as

	YoGi	YoBo	OlGi	OlBo
SC	10.46	7.64	10.21	8.11
SA	5.13	8.36	9.53	0
AC	9.12	3.02	0	8.47
PA	6.73	7.06	6.51	6.24
BC	7.37	7.68	4.40	8.38
GS	0.63	0	0	0

,

where the unique variance of zero is a boundary solution. This may also occur for other models and estimation methods; see e.g. Bentler and Lee (1979).

The percentage of explained common variance for each k is: 88.95, 85.48, 86.03, 81.53. The percentage of explained common variance for each variable j and each sample k, $j = 1, \ldots, 6$, $k = 1, \ldots, 4$ is given as follows

	SC	SA	AC	PA	BC	GS
YoGi	81.53	85.29	78.81	94.06	73.03	97.32
YoBo	77.56	81.92	86.39	92.14	70.61	96.13
OlGi	75.13	87.05	77.11	95.05	62.71	98.13
OlBo	53.11	83.21	34.09	98.59	72.63	97.16

.

The percentage of explained common variance due to each factor for each sample is given as

	factor1	factor2
YoGi	68.36	20.59
YoBo	49.81	35.66
OlGi	67.95	18.08
OlBo	59.61	21.92

.

The solution (6.18) can be interpreted as follows. Factor 1 is a general factor with highest loadings for PA and GS. Factor 2 is a combination of SC, SA, and AC. The weights $\mathbf{C}_k$ shows that there are differences in factors between groups. Groups YoGi and OlGi have larger weights than groups YoBo and OlBo for factor 1, whereas group YoBo has larger weight than the other groups for factor 2. For groups YoGi, OlGi, and OlBo, factor 1 has a considerably larger weight than factor 2. For group YoBo, both factors have almost the same weight. The percentages of explained common variances for group OlBo are rather low for the sumscores SC and AC. Compared to other groups, the percentage of explained common variance of YoBo is smallest for factor 1 but largest for factor 2. The total percentage of explained common variance for groups YoBo and OlBo together is smaller than the percentage for groups YoGi and OlGi together.

In general, the results show that there are differences between boys and girls, and between young and old children. Factor 1 represents Physical Appearance (PA) and Global Self-worth (GS) and is stronger for the girls than for the boys. Hence, factor 1 captures differences in the variability of PA and GS judgements between boys and girls. Factor 2 represents Scholastic Competence, Social Acceptance, and Athletic Competence, and is much stronger for young boys than for the other groups. The young boys apparently show more variability on these scales than the other groups. This may be due to their lack of a coherent self-perception.

For $R = 2$ oblique factors the solution is very similar to $R = 2$ orthogonal factors, with factor correlation -0.07. Note that the solutions are essentially unique due to the uniqueness condition of Ten Berge and Kiers (1996).

We also fit the model with $R = 3$ orthogonal factors. In the solution, the first and third factors are similar to the two factors in (6.18), and the second

factor is a combination of AC, PA and GS. The percentage of explained common variance for group YoBo is also smallest on factor 1 and largest on factor 3. The interpretation of this solution is similar to the $R = 2$ solution, with the addition of a new factor. To save space, we do not present the $R = 3$ solution in detail. Finally, we remark that the $R = 3$ solution is the same for many runs of our algorithm in section 6.3.2. Hence, it seems to be essentially unique.

6.6 Discussion

In this chapter, we have proposed a new method to fit the indirect Parafac2 model via the direct Parafac2 model, and presented an exploratory multi-set factor model with common covariance part of indirect Parafac2 form. To estimate our multi-set Parafac2 factor model we use MRFA to obtain the unique variances $\mathbf{U}_k$ and our indirect Parafac2 algorithm to estimate the common loading matrix $\mathbf{B}$, factor strengths $\mathbf{C}_k$ for each sample k, and the factor correlation matrix $\mathbf{\Phi}$. The matrices $\mathbf{\Sigma}_k - \mathbf{U}_k$ are guaranteed to be covariance matrices due to the MRFA algorithm. Therefore, percentages of explained common variance can be computed for each sample k, and for each variable in each sample k. For other methods of multi-set data analysis, such as simultaneous component analysis (SCA) or multigroup exploratory (EFA) or confirmatory (CFA) factor analysis, it is not guaranteed that such explained common variances can be computed. Our exploratory multi-set Parafac2 factor model relates to the SCA models as EFA relates to PCA for one dataset. Also, our exploratory multi-set Parafac2 factor model relates to multigroup CFA as EFA relates to CFA for one dataset.

The simulations in section 6.2.3 show that our indirect Parafac2 algorithm performs better than that of Kiers (1993) on random datasets with perfect fit.

For noisy data, no comparison has been made to Kiers (1993), but the simulation study in section 6.2.4 shows that our indirect Parafac2 algorithm performs well. The simulation study in section 6.4 shows that our relatively simple estimation procedure for the multi-set Parafac2 factor model in section 6.3.2 performs very well in retrieving underlying factors when the data is randomly sampled with true covariance matrices $\mathbf{\Sigma}_k - \mathbf{U}_k$ satisfying the indirect Parafac2 model. The recovery is better when we use orthogonal factors.

The results of application in section 6.5 confirm the considerations of Meijer et al. (2008) on differences between young children and old children, and between girls and boys when they judge their own functioning in several specific domains and their global self-worth. Our results show that girls have higher variability in their judgement on Physical appearance and Global Self-worth than boys. Also, young boys have higher variability in their judgements on Scholastic Competence, Social Acceptance, and Athletic Competence.

Parafac2 solutions are essentially unique under mild conditions for $R = 2$ and $K \geq 4$ due to the result of Ten Berge and Kiers (1996). Based on simulations, Kiers et al. (1999) suggest that Parafac2 uniqueness holds usually for $K \geq 4$ and any $R \geq 2$. Such Parafac2 uniqueness results remain to be proven, however.

Appendix: SPPC covariance matrices of the sumscore on the six subscales

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 14.38 & 4.97 & 3.02 & 2.20 & 3.29 & 4.31 \\ 4.97 & 14.90 & 5.40 & 4.40 & 3.35 & 6.60 \\ 3.02 & 5.40 & 13.50 & 4.44 & 3.26 & 4.66 \\ 2.20 & 4.40 & 4.44 & 20.11 & 5.29 & 12.18 \\ 3.29 & 3.35 & 3.26 & 5.29 & 11.80 & 6.11 \\ 4.31 & 6.60 & 4.66 & 12.18 & 6.11 & 13.91 \end{pmatrix}, \quad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 17.68 & 5.77 & 6.89 & 2.27 & 5.94 & 4.38 \\ 5.77 & 15.17 & 6.24 & 4.68 & 4.24 & 6.20 \\ 6.89 & 6.24 & 12.51 & 2.46 & 3.52 & 2.56 \\ 2.27 & 4.68 & 2.46 & 14.80 & 3.16 & 8.11 \\ 5.94 & 4.24 & 3.52 & 3.16 & 12.31 & 4.36 \\ 4.38 & 6.20 & 2.56 & 8.11 & 4.36 & 10.52 \end{pmatrix},$$

$$\boldsymbol{\Sigma}_3 = \begin{pmatrix} 13.82 & 3.88 & 4.37 & 3.91 & 3.59 & 4.52 \\ 3.88 & 15.23 & 6.15 & 7.72 & 4.58 & 7.70 \\ 4.37 & 6.15 & 13.36 & 7.30 & 2.60 & 5.91 \\ 3.91 & 7.72 & 7.30 & 19.62 & 5.64 & 12.49 \\ 3.59 & 4.58 & 2.60 & 5.64 & 9.65 & 6.09 \\ 4.52 & 7.70 & 5.91 & 12.49 & 6.09 & 12.95 \end{pmatrix}, \quad \boldsymbol{\Sigma}_4 = \begin{pmatrix} 12.64 & 4.45 & 0.77 & 3.46 & 3.73 & 2.79 \\ 4.45 & 12.42 & 4.72 & 4.66 & 3.42 & 4.68 \\ 0.77 & 4.72 & 10.78 & 2.13 & 0.81 & 2.37 \\ 3.46 & 4.66 & 2.13 & 13.61 & 4.62 & 7.51 \\ 3.73 & 3.42 & 0.81 & 4.62 & 12.33 & 4.45 \\ 2.79 & 4.68 & 2.37 & 7.51 & 4.45 & 8.10 \end{pmatrix}.$$

Chapter 7

Conclusion

In this concluding chapter we classify the types of models discussed in this thesis, provide a brief discussion of model choice, and consider options for future research.

The component and factor models for three-mode and multi-set data (new and existing) that have been discussed in this thesis can be arranged in a $2 \times 2 \times 3$ classification array. The first mode of the array represents the three-mode/multi-set dichotomy. The second mode indicates whether the model is rotationally unique or not. The third mode represents the type of model: direct-fitting, indirect-fitting, or a factor model. Direct-fitting models are fitted on the observed data itself, which is given by $\mathbf{X}_{(N\times JK)}$ for three-way data, and $\mathbf{X}_k$, $k = 1, \ldots, K$, for multi-set data. Indirect-fitting models are fitted to the observed cross-products of the data, which we divide by the number of observations N to make a comparison with the factor models. Thus, the cross-products are given by $\mathbf{\Sigma} = N^{-1}\mathbf{X}_{(N\times JK)}^T\mathbf{X}_{(N\times JK)}$ for three-way data, and $\mathbf{\Sigma}_k = N^{-1}\mathbf{X}_k^T\mathbf{X}_k$, $k = 1, \ldots, K$ for multi-set data. For the factor models, we assume mean-zero data and fit the model to the above cross-products, which are now covariance

matrices. Additionally, we assume a unique part that is uncorrelated with the factors for the common part.

Table 7.1: Classification of three-mode and multi-set component and factor models discussed in this thesis.

three-mode rotationally unique	$\mathbf{X}_{(N\times JK)} = \mathbf{F}\,(\mathbf{C}\odot\mathbf{B})^T$ $\mathbf{\Sigma} = (\mathbf{C}\odot\mathbf{B})\,\mathbf{\Phi}\,(\mathbf{C}\odot\mathbf{B})^T\ (+\mathbf{U})$
three-mode rotationally nonunique	$\mathbf{X}_{(N\times JK)} = \mathbf{F}\,\mathbf{G}\,(\mathbf{C}\otimes\mathbf{B})^T$ $\mathbf{\Sigma} = (\mathbf{C}\otimes\mathbf{B})\,\mathbf{G}^T\,\mathbf{\Phi}\,\mathbf{G}\,(\mathbf{C}\otimes\mathbf{B})^T\ (+\mathbf{U})$
multi-set rotationally unique	$\mathbf{X}_k = \mathbf{F}_k\,\mathbf{C}_k\,\mathbf{B}^T, \quad \mathbf{F}_k^T\mathbf{F}_k = \mathbf{F}_l^T\mathbf{F}_l$ $\mathbf{\Sigma}_k = \mathbf{B}\,\mathbf{C}_k\,\mathbf{\Phi}\,\mathbf{C}_k\,\mathbf{B}^T\ (+\mathbf{U}_k)$
multi-set rotationally nonunique	$\mathbf{X}_k = \mathbf{F}_k\,\left(\sum_{q=1}^{Q} c_{kq}\mathbf{G}_q\right)\,\mathbf{B}^T, \quad \mathbf{F}_k^T\mathbf{F}_k = \mathbf{F}_l^T\mathbf{F}_l$ $\mathbf{\Sigma}_k = \mathbf{B}\,\left(\sum_{q=1}^{Q} c_{kq}\mathbf{G}_q\right)^T\,\mathbf{\Phi}\,\left(\sum_{q=1}^{Q} c_{kq}\mathbf{G}_q\right)\,\mathbf{B}^T\ (+\mathbf{U}_k)$

In Table 7.1 above, the models are given for each combination of the first two modes of classification. In our classification, rotationally unique three-mode models are of CP form, rotationally nonunique three-mode models are of the Tucker form, rotationally unique multi-set models are of Parafac2 form, and rotationally nonunique multi-set models are of multi-set Tucker form. The last model type may be new in the literature and is not discussed in this thesis. The three-mode factor models in Table 7.1 are easy to extend to more than three modes. The indirect and factor multi-set models are obtained from the corresponding three-mode models by taking the kth $J \times J$ diagonal block of the $JK \times JK$ matrix $\mathbf{\Sigma}$. Note that Table 7.1 is not meant as a classification of all possible or existing models of these types. For example, there also exist other

types of multi-set component models (Timmerman & Kiers, 2003).

The choice of model obviously depends on the data type. For three-mode data, we can choose between three-mode models or multi-set models. It may seem obvious to prefer three-mode models in this case, but multi-set models also have advantages when fitted to three-mode data (De Roover, Timmerman, Van Mechelen, & Ceulemans, 2013). For the choice between a rotationally unique or nonunique model, we recommend to try both. In our classification, the rotationally unique models (CP-based) are a special case of rotationally nonunique models (Tucker-based). After rotating, the Tucker-based solutions can be compared to solutions from CP-based models. As we have seen in Chapter 5, for one three-mode dataset the CP-based factor model may be more suitable, while for another three-mode dataset the Tucker-based factor model yields a clearer interpretation. Naturally, the Tucker-based models involve more elaborate choices regarding numbers of components and rotation methods.

The choice between a component model (direct or indirect) or a factor model is a fundamental one. Component models aim at a best-fitting summary of the data as it is observed. The components are found as linear combinations of the observed variables. Factor models aim at taking into account measurement error and aspects measured by single observed variables only, by assuming a unique part for each observed variable. A summary of the common part of the observed variables is then sought in terms of the common factors, where the latter are usually not linear combinations of the observed variables. For two-way data, a lively discussion of PCA versus (exploratory) common factor analysis (EFA) can be found in Velicer and Jackson (1990). PCA can be favored because of its computational simplicity and manifest component approach, whereas EFA is computationally more difficult and features latent factors. PCA and EFA are said to yield similar estimated loadings in simulation studies. However, the

PCA loadings are difficult to compare for different samples, since no unique parts have been specified. Apart from the results from simulation studies, the same arguments apply to the three-mode and muti-set component and factor models in Table 7.1.

The factor models in this thesis are more difficult to estimate than the corresponding component models. We have proposed a two-step estimation procedure for the factor models, where we first use MRFA to estimate the unique variances. Next, the covariance matrix (or matrices) of the common part(s) is decomposed and the corresponding direct-fitting component model is used to estimate the factor loadings. Using MRFA has as advantage that percentages of explained common variance can be computed, overall and for each observed variable separately. A disadvantage may be that MRFA imposes an additional constraint ($\mathbf{\Sigma} - \mathbf{U} \geq 0$) on the factor solution, which is assumed to be true in the population but may not hold in the sample at hand. Also, our use of MRFA is not optimal since the form of the model for the common part is not taken into account when estimating the unique variances. However, despite these potential limitations, the simulation studies in Chapter 4 and Chapter 6 have shown that our rotationally unique factor models and estimation methods are successful in retrieving the true underlying factor loadings.

Finally, we indicate some topics for future research. As in Chapters 4 and 6, a simulation study could be done to assess the performance of the estimation method for the Tucker3 three-mode factor model in Chapter 5. This involves choosing a rotation method to rotate the true and estimated solutions to simple structure. Also, it would be interesting to apply our approach to the multi-set Tucker3 factor model (the final model in Table 7.1). For this, we propose an analogous estimation procedure as for the multi-set Parafac2 factor model in section 6.3.2. In step 3 of this procedure, the direct-fitting multi-set Tucker3

model should be fitted. This can be done analogous to the direct Parafac2 algorithm in section 6.2.1, by replacing the fitting of CP in step 2 of this procedure by the fitting of Tucker3. As for the other factor models, a simulation study can be conducted and the model should be fitted to multi-set data. For example, it would be interesting to fit the multi-set Tucker3 factor model to the SPPC dataset in Chapter 6 and compare the solution(s) to the solution of the multi-set Parafac2 factor model.

Another topic for future research is to conduct simulation studies to compare the performance of the factor models in Table 7.1 to that of the corresponding indirect component models, i.e., without a unique part. As mentioned above, of PCA and EFA it is said that they yield similar loading estimates. Would this be true for the three-mode and multi-set component and factor models as well? Or would the suboptimal two-step approach hamper the performance of the estimation procedures of the factor models, as compared to fitting the corresponding component model? Or would the performance rather be hampered by the use of MRFA instead of a less restrictive MINRES approach? These are interesting research directions to pursue. In our opinion, however, the choice between a component model and a factor model remains a fundamental one that should not be based solely on estimation accuracy.

Bibliography

Acar, E., & Yener, B. (2009). Unsupervised multiway data analysis: a literature survey. *IEEE Transactions on Knowledge and Data Engineering*, *21*(1), 6–20. doi: http://doi.ieeecomputersociety.org/10.1109/TKDE.2008.112

Bentler, P. M., & Lee, S. Y. (1978). Statistical aspects of a three-mode factor analysis model. *Psychometrika*, *43*(3), 343–352. doi: 10.1007/BF02293644

Bentler, P. M., & Lee, S. Y. (1979). A statistical development of three-mode factor analysis. *British Journal of Mathematical and Statistical Psychology*, *32*(1), 87–104. doi: 10.1111/j.2044-8317.1979.tb00754.x

Bloxom, B. (1968). A note on invariance in three-mode factor analysis. *Psychometrika*, *33*(3), 347-350. doi: 10.1007/BF02289329

Bro, R., Harshman, R. A., Sidiropoulos, N. D., & Lundy, M. E. (2009). Modeling multi-way data with linearly dependent loadings. *Journal of Chemometrics*, *23*(7-8), 324–340. doi: 10.1002/cem.1206

Browne, M. W. (1984). The decomposition of multitrait-multimethod matrices. *British Journal of Mathematical and Statistical Psychology*, *37*(1), 1–21. doi: 10.1111/j.2044-8317.1984.tb00785.x

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*(1), 111–150. doi: 10.1207/S15327906MBR3601_05

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. doi: 10.1037/h0046016

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an n-way generalization of Eckart-Young de-

composition. *Psychometrika*, *35*(3), 283–319. doi: 10.1007/BF02310791

Cattell, R. B. (1944). Parallel proportional profiles and other principles for determining the choice of factors by rotation. *Psychometrika*, *9*(4), 267–283. doi: 10.1007/BF02288739

Ceulemans, E., & Kiers, H. A. L. (2006). Selecting among three-mode principal component models of different types and complexities: a numerical convex hull based method. *British Journal of Mathematical and Statistical Psychology*, *59*(1), 133–150. doi: 10.1348/000711005X64817

Comon, P., & De Lathauwer, L. (2010). Algebraic identification of underdetermined mixtures. In P. Comon & C. Jutten (Eds.), *Handbook of blind source separation: Independent component analysis and applications* (pp. 325–366). Academic Press. doi: http://dx.doi.org/10.1016/B978-0-12-374726-6.00014-X

Comon, P., Luciani, X., & De Almeida, L. F. (2009). Tensor decompositions, alternating least squares and other tales. *Journal of Chemometrics*, *23*(7-8), 393–405. doi: 10.1002/cem.1236

Crawford, C. B., & Ferguson, G. A. (1970). A general rotation criterion and its use in orthogonal rotation. *Psychometrika*, *35*(3), 321–332. doi: 10.1007/BF02310792

De Almeida, A. L. F., Favier, G., & Mota, J. C. M. (2008a). A constrained factor decomposition with application to MIMO antenna systems. *IEEE Transactions on Signal Processing*, *56*(6), 2429–2442. doi: 10.1109/TSP.2008.917026

De Almeida, A. L. F., Favier, G., & Mota, J. C. M. (2008b). Constrained tensor modeling approach to blind multiple-antenna CDMA schemes. *IEEE Transactions on Signal Processing*, *56*(6), 2417–2428. doi: 10.1109/TSP.2008.917028

De Lathauwer, L. (2008). Decompositions of a higher-order tensor in block terms. Part II: Definitions and Uniqueness. *SIAM Journal on Matrix Analysis and Applications*, *30*(3), 1033–1066. doi: 10.1137/070690729

De Lathauwer, L. (2010). Algebraic methods after prewhitening. In P. Comon & C. Jutten (Eds.), *Handbook of blind source separation: Independent component analysis and applications* (pp. 155–177). Academic Press. doi: http://dx.doi.org/10.1016/B978-0-12-374726-6.00010-2

De Lathauwer, L., De Moor, B., & Vandewalle, J. (2000). A multilinear singular value decomposition. *SIAM Journal on Matrix Analysis and Applications*, *21*(4), 1253–1278. doi: 10.1137/S0895479896305696

De Roover, K., Timmerman, M. E., Van Mechelen, I., & Ceulemans, E. (2013). On the added value of multiset methods for three-way data analysis. *Chemometrics and Intelligent Laboratory Systems*, *129*, 98–107. doi: 10.1016/j.chemolab.2013.05.002

De Silva, V., & Lim, L. H. (2008). Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal on Matrix Analysis and Applications*, *30*(3), 1084–1127. doi: 10.1137/06066518X

Dickinson, T. L., & Tice, T. E. (1973). A multitrait-multimethod analysis of scales developed by retranslation. *Organizational Behavior and Human Performance*, *9*(3), 421–438. doi: http://dx.doi.org/10.1016/0030-5073(73)90063-9

Domanov, I., & De Lathauwer, L. (2013a). On the Uniqueness of the Canonical Polyadic Decomposition of Third-Order Tensors-Part I: Basic Results and Uniqueness of One Factor Matrix. *SIAM Journal on Matrix Analysis and Applications*, *34*, 855-875.

Domanov, I., & De Lathauwer, L. (2013b). On the Uniqueness of the Canonical Polyadic Decomposition of Third-Order Tensors-Part II: Uniqueness of the

Overall Decomposition. *SIAM Journal on Matrix Analysis and Applications*, *34*, 876-903.

Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, *1*(3), 211–218. doi: 10.1007/BF02288367

Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, *65*(2), 241–261. doi: 10.1007/BF02294377

Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: different models for different types of methods. *Psychological Methods*, *13*(3), 230–253. doi: 10.1037/a0013219

Harman, H. H., & Jones, W. H. (1966). Factor analysis by minimizing residuals (MINRES). *Psychometrika*, *31*(3), 351–368. doi: 10.1007/BF02289468

Harshman, R. A. (1970). Foundations of the Parafac proceduce: models and conditions for an "explanatory" multimodal factor analysis. *UCLA Working papers in Phonetics*, *16*, 1–84.

Harshman, R. A. (1972). Parafac2: Mathematical and technical notes. *UCLA Working Papers in Phonetics*, *22*, 30–44.

Harshman, R. A. (2004). The problem and nature of degenerate solutions or decompositions of 3-way arrays. In *Tensor decompositions workshop*, Palo Alto, USA.

Harshman, R. A., & Lundy, M. E. (1984). Data preprocessing and the extended Parafac model. In H. G. Law, C. W. Snyder Jr., J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 216–284). Praeger, New York.

Harshman, R. A., & Lundy, M. E. (1996). Uniqueness proof for a family of model sharing features of Tucker's three-mode factor analysis and Parafac/Candecomp. *Psychimetrika*, *61*(1), 133–154. doi: 10.1007/

BF02296963

Hitchcock, F. L. (1927a). The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, *6*, 164–189.

Hitchcock, F. L. (1927b). Multiple invariants and generalized rank of a p-way matrix or tensor. *Journal of Mathematics and Physics*, *7*, 39–70.

Holzinger, K., & Swineford, F. (1939). *A study in factor analysis: The stability of a bi-factor solution.* Chicago: University of Chicago Press.

Ishteva, M., Absil, P. A., Van Huffel, S., & De Lathauwer, L. (2011). Best low multilinear rank approximation of higher-order tensors, based on the Riemannian trust-region scheme. *SIAM Journal on Matrix Analysis and Applications*, *32*(1), 115–135. doi: 10.1137/090764827

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*(2), 183–202. doi: 10.1007/BF02289343

Jöreskog, K. G. (1970). A general method for analysis of covariance structures. *Biometrika*, *57*(2), 239–251. doi: 10.2307/2334833

Jöreskog, K. G. (1971a). Simultaneous factor analysis in several populations. *Psychometrika*, *36*(4), 409–426. doi: 10.1007/BF02291366

Jöreskog, K. G. (1971b). Statistical analysis of sets of congeneric tests. *Psychometrika*, *36*(2), 109–133. doi: 10.1007/BF02291393

Jennrich, R. I. (1970). Orthogonal rotation algorithms. *Psychometrika*, *35*(2), 229–235. doi: 10.1007/BF02291264

Jiang, T., & Sidiropoulos, N. D. (2004). Kruskal's permutation lemma and the identification of Candecomp/Parafac and bilinear models with constant modulus constraints. *IEEE Transactions on Signal Processing*, *52*(9), 2625 – 2636. doi: 10.1109/TSP.2004.832022

Kiers, H. A. L. (1993). An alternating least squares algorithm for PARAFAC2

and three-way Dedicom. *Computational Statistics & Data Analysis*, *16*(1), 103–118. doi: http://dx.doi.org/10.1016/0167-9473(93)90247-Q

Kiers, H. A. L. (1998a). Joint Orthomax rotation of the core and component matrices resulting from three-mode Principal Component Analysis. *Journal of Classification*, *15*(2), 245–263. doi: 10.1007/s003579900033

Kiers, H. A. L. (1998b). Three-way SIMPLIMAX for oblique rotation of the three-mode factor analysis core to simple structure. *Computational Statistics and Data Analysis*, *28*(3), 307–324. doi: http://dx.doi.org/10.1016/S0167-9473(98)00041-3

Kiers, H. A. L., Kroonenberg, P. M., & Ten Berge, J. M. F. (1992). An efficient algorithm for TUCKALS3 on data with large number of observation units. *Psychometrika*, *57*(3), 415–422. doi: 10.1007/BF02295429

Kiers, H. A. L., Takane, Y., & Ten Berge, J. M. F. (1996). The analysis of multitrait-multimethod matrices via constrained component analysis. *Psychometrika*, *61*(4), 601–628. doi: 10.1007/BF02294039

Kiers, H. A. L., Ten Berge, J. M. F., & Bro, R. (1999). Parafac2-Part I: A direct fitting algorithm for the Parafac2 model. *Journal of Chemometrics*, *13*(3-4), 601–628. doi: 10.1002/(SICI)1099-128X(199905/08)13:3/4⟨275::AID-CEM543⟩3.0.CO;2-B

Kiers, H. A. L., & Van Mechelen, I. (2001). Three-way component analysis: principles and illustrative application. *Psychological Methods*, *6*(1), 84–110. doi: 10.1037//1082-989X.6.1.84

Kolda, T. G., & Bader, B. W. (2009). Tensor decompositions and applications. *SIAM Review*, *51*(3), 455–500. doi: http://dx.doi.org/10.1137/07070111X

Krijnen, W. P., Dijkstra, T. K., & Stegeman, A. (2008). On the non-exsistence of optimal solutions and the occurence of "degeneracy" in

the Candecomp/Parafac model. *Psychometrika*, *73*(3), 431–439. doi: 10.1007/s11336-008-9056-1

Kroonenberg, P. M. (2008). *Applied Multiway Data Analysis.* Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley.

Kroonenberg, P. M., & De Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, *45*(1), 69–97. doi: 10.1007/BF02293599

Kroonenberg, P. M., & Oort, F. J. (2003). Three-mode analysis of multimode covariance matrices. *British Journal of Mathematical and Statistical Psychology*, *56*(2), 305–335. doi: 10.1348/000711003770480066

Kruskal, J. B. (1977). Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics. *Linear Algebra and its Applications*, *18*(2), 95–138. doi: http://dx.doi.org/10.1016/0024-3795(77)90069-6

Kruskal, J. B. (1989). Rank, Decomposition, and Uniqueness for three-way and N-way arrays. In R. Coppi & S. Bolassco (Eds.), *Multiway data analysis.* Elsevier Science Publisher B.V. (North-Holland).

Kruskal, J. B., Harshman, R. A., & Lundy, M. E. (1989). How 3-MFA data can cause degenerate Parafac solutions, among other relationships. In R. Coppi & S. Bolassco (Eds.), *Multiway data analysis* (pp. 115–121). Elsevier Science Publisher B.V. (North-Holland).

Lam, T. T. T., & Stegeman, A. (2014). Multi-set fator analysis by means of Parafac2. *British Journal of Mathematical and Statistical Psychology*. (Submitted)

Lerner, M. J. (1980). *The belief in a just world: A fundamental delusion.* New York: Plenum.

Lim, L. H., & Comon, P. (2009). Nonnegative approximations of nonnegative

tensors. *Journal of Chemometrics*, *23*(7-8), 432–441. doi: 10.1002/cem.1244

Lipkus, I., Dalbert, C., & Siegler, I. C. (1996). The importance of distinguishing the belief in a just world for self versus for others: Implications for psychological well-being. *Personality and Social Psychology Bulletin*, *22*(7), 666–677. doi: 10.1177/0146167296227002

Liu, X., & Sidiropoulos, N. D. (2001). Cramér-Rao lower bounds for low-rank decomposition of multidimensioanl arrays. *IEEE Transactions on Signal Processing*, *49*(9), 2074–2086. doi: 10.1109/78.942635

Lorenzo-Seva, U., & Ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *2*(2), 57–64. doi: 10.1027/1614-2241.2.2.57

Meijer, R. R., Egberink, I. J. L., Emons, W. H. M., & Sijtsma, K. (2008). Detection and validation of unscalable item score patterns using item response theory: an illustration with Harter's Self-perception profile for children. *Journal of Personality Assessment*, *90*(3), 227–238. doi: 10.1080/00223890701884921

Millsap, R. E. (1992). Sufficient conditions for rotational uniqueness in the additive MTMM model. *British Journal of Mathematical and Statistical Psychology*, *45*(1), 125–138. doi: 10.1111/j.2044-8317.1992.tb00982.x

Ogasawara, H. (2000). Some relationships between factors and components. *Psychometrika*, *65*(2), 167–185. doi: 10.1007/BF02294372

Paatero, P. (2000). Construction and analysis of degenerate PARAFAC models. *Journal of Chemometrics*, *14*(3), 285–299. doi: 10.1002/1099-128X(200005/06)14:3⟨285::AID-CEM584⟩3.0.CO;2-1

Pearson, K. (1901). On lines and planes of closest fit to systems of points in

space. *Philosophical Magazine*, *2*, 559–572.

Rhodes, J. A. (2010). A concise proof of Kruskal's theorem on tensor decomposition. *Linear Algebra and its Applications*, *432*(7), 1818 – 1824. doi: http://dx.doi.org/10.1016/j.laa.2009.11.033

Rocci, R., & Giordani, P. (2010). A weak degeneracy revealing decomposition for the CANDECOMP/PARAFAC model. *Journal of Chemometrics*, *24*(2), 57–66. doi: 10.1002/cem.1272

Savas, B., & Lim, L. H. (2010). Quasi-Newton methods on Grassmannians and multilinear approximation of tensors. *SIAM Journal on Scientific Computing*, *32*(6). doi: 10.1137/090763172

Smilde, A., Bro, R., & Geladi, P. (2004). *Multi-way Analysis: Applications in the Chemical Sciences*. Chichester: Wiley.

Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, *27*(2), 229–239. doi: 10.1111/j.2044-8317.1974.tb00543.x

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.

Stegeman, A. (2006). Degeneracy in Candecomp/Parafac explained for $p \times p \times 2$ arrays of rank $p + 1$ or higher. *Psychometrika*, *71*(3), 483–501. doi: 10.1007/s11336-004-1266-6

Stegeman, A. (2007). Degeneracy in Candecomp/Parafac and Indscal explained for several three-sliced arrays with a two-valued typical rank. *Psychometrika*, *72*(4), 601–619. doi: 10.1007/s11336-007-9022-3

Stegeman, A. (2008). Low-rank approximation of generic $p \times q \times 2$ arrays and diverging components in the Candecomp/Parafac model. *SIAM Journal on Matrix Analysis and Applications*, *30*(3), 988–1007. doi: 10.1137/

050644677

Stegeman, A. (2009a). On uniqueness conditions for Candecomp/Parafac and Indscal with full column rank in one mode. *Linear Algebra and its Applications*, *431*(1-2), 211–227. doi: http://dx.doi.org/10.1016/j.laa.2009.02.025

Stegeman, A. (2009b). Using the Simultaneous Generalized Schur Decomposition as a Candecomp/Parafac algorithm for ill-conditioned data. *Journal of Chemometrics*, *23*(7-8), 385–392. doi: 10.1002/cem.1232

Stegeman, A. (2010). The Generalized Schur Decomposition and the rank-R set of real $I \times J \times 2$ arrays. *Technical Report*. (Available online as arXiv: 1011.3432)

Stegeman, A. (2011). On uniqueness of the canonical tensor decomposition with some form of symmetry. *SIAM Journal on Matrix Analysis and Applications*, *32*(2), 561–583. doi: 10.1137/100814615

Stegeman, A. (2012). Candecomp/Parafac - from diverging components to a decomposition in block terms. *SIAM Journal on Matrix Analysis and Applications*, *33*(2), 291–316. doi: 10.1137/110825327

Stegeman, A. (2013). A three-way jordan canonical form as limit of low-rank tensor approximations. *SIAM Journal on Matrix Analysis and Applications*, *34*(2), 624–650. doi: 10.1137/120875806

Stegeman, A. (2014). Finding the limit of diverging components in three-way Candecomp/Parafac - a demonstration of its practical merits. *Computational Statistics & Data Analysis*, *75*. doi: http://dx.doi.org/10.1016/j.csda.2014.02.010

Stegeman, A., & Almeida, A. L. F. (2009). Uniqueness conditions for constrained three-way factor decompositions with linearly dependent loadings. *SIAM Journal on Matrix Analysis and Applications*, *31*(3), 1469–1490. doi: 10

.1137/080743354

Stegeman, A., & De Lathauwer, L. (2009). A method to avoid diverging components in the Candecomp/Parafac model for generic $I \times J \times 2$ arrays. *SIAM Journal on Matrix Analysis and Applications*, *30*(3), 1614–1638. doi: 10.1137/080743354

Stegeman, A., & Lam, T. T. T. (2012). Improved uniqueness conditions for canonical tensor decompositions with linearly dependent loadings. *SIAM Journal on Matrix Analysis and Applications*, *33*(4), 1250–1271. doi: 10 .1137/110847275

Stegeman, A., & Lam, T. T. T. (2014). Three-mode factor analysis by means of Candecomp/Parafac. *Psychometrika*, *79*(3), 426–443. doi: 10.1007/ S11336-013-9359-8

Stegeman, A., & Sidiropoulos, N. D. (2007). On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition. *Linear Algebra and its Applications*, *420*(2-3), 540 – 552. doi: http://dx.doi.org/10.1016/ j.laa.2006.08.010

Stegeman, A., Ten Berge, J. M. F., & De Lathauwer, L. (2006). Sufficient conditions for uniqueness in Candecomp/Parafac and Indscal with random component matrices. *Psychometrika*, *71*(2), 219–229. doi: 10.1007/11336 -006-1278-2

Stroebe, K., Postmes, T., Täuber, S., Stegeman, A., & John, M. S. (2014). Belief in a just what? Demystifying just world belief by distinguishing sources of justice. *Submitted*.

Sŏcan, G. (2003). *The Incremental Value of Minimum Rank Factor Analysis* (PhD thesis). Department of Psychometrics & Statistics, University of Groningen, Groningen, The Netherlands.

Ten Berge, J. M. F. (2011). Simplicity and typical rank results for three-way

arrays. *Psychometrika*, *76*(1), 3–12. doi: 10.1007/s11336-010-9193-1

Ten Berge, J. M. F., & Kiers, H. A. L. (1991). A numerical approach to the approximate and the exact minimum rank of a covariance matrix. *Psychometrika*, *56*(2), 309–315. doi: 10.1007/BF02294464

Ten Berge, J. M. F., & Kiers, H. A. L. (1996). Some uniqueness results for Parafac2. *Psychometrika*, *61*(1), 123 – 132. doi: 10.1007/BF02296962

Ten Berge, J. M. F., & Sidiropoulos, N. D. (2002). On uniqueness in Candecomp/Parafac. *Psychometrika*, *67*(3), 399–409. doi: 10.1007/BF02294992

Ten Berge, J. M. F., & Stegeman, A. (2006). Symmetry transformations for square sliced three-way arrays, with applications to their typical rank. *Linear Algebra and its Applications*, *418*(1), 215 – 224. doi: http://dx.doi.org/10.1016/j.laa.2006.02.002

Tendeiro, J. N., Ten Berge, J. M. F., & Kiers, H. A. L. (2009). Simplicity transformations for three-way arrays with symmetric slices, and applications to Tucker-3 models with sparse core arrays. *Linear Algebra and its Applications*, *430*(4), 924–940. doi: http://dx.doi.org/10.1016/j.laa.2008.09.020

Thurstone, L. L. (1935). *The vectors of mind*. Chicago: University of Chicago Press.

Timmerman, M., & Kiers, H. A. L. (2003). Four Simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences. *Psychometrika*, *68*(1), 105–121. doi: 10.1007/BF02296656

Tipping, M. E., & Bishop, C. M. (1999). Mixtures of probabilistic principal component analyzers. *Neural Computation*, *11*(2), 443–482. doi: 10.1162/089976699300016728

Tomasi, G., & Bro, R. (2006). A comparison of algorithms for fitting the Parafac

model. *Computational Statistics and Data Analysis*, *50*(7), 1700–1734. doi: http://dx.doi.org/10.1016/j.csda.2004.11.013

Tucker, L. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, *31*(3), 279–311. doi: 10.1007/BF02289464

Velicer, W. F., & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure (with comments and reply). *Multivariate Behavioral Research*, *25*(1), 1–114. doi: 10.1207/s15327906mbr2501_1

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, *9*(1), 1–26. doi: 10.1177/014662168500900101

Wothke, W. (1996). Models for multitrait-multimethod matrix analysis. In G. Marcoulides & R. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 7–56). Mahwah, NJ:Erlbaum.

Summary

This thesis discussed methods for Exploratory Component and Factor Analysis of multi-way and multi-set data. In particular, new methods for multi-way and multi-set factor analysis were proposed and demonstrated. Following is a summary of the contents of each chapter of the thesis.

In Chapter 2 we introduced the concepts of multi-way and multi-set component and factor models whose comprehension was crucial to fully understand all the subsequent chapters of this thesis. These concepts concern two-way decompositions (e.g., Singular Valued Decomposition, Principal Component Analysis), three-way decompositions (e.g., Candecomp/Parafac, Tucker3), multi-set component and factor analysis (e.g., Parafac2, Simultaneous Component Analysis models, Multi-set Parafac2 common factor model). Additionally, we discussed the uniqueness properties of both Candecomp/Parafac and Tucker3 models, and algorithms to fit these models.

Chapter 3 was devoted to multi-way exploratory analysis of 4-way BJW data in which the subjects are asked how strongly they believe that a number of 6 actors (Nature, God, Human Institutions, Other People, Yourself, and Chance) bring about justice in the world for themselves or other people. The 4 modes of the dataset are: 345 subjects, 7 items, 6 actors, and 2 perspectives (Yourself and Others). In this chapter, we included the results of preliminary analysis of correlations and t-tests, and of PCA on the BJW dataset. Moreover, we also fitted the Tucker3 and Tucker4 models to this dataset. Compared to PCA on unfolding data, the Tucker3 and Tucker4 solutions are almost the same. They also have simple structures and easy interpretation. However, they are a more parsimonious summary than PCA solutions of 2-way unfoldings of the dataset. The Tucker4 solution is the most parsimonious summary because it is

the solution of the complete dataset. The Tucker4 rotation was discussed at the end of this chapter.

Chapters 4, 5, 6 proposed and demonstrated new methods for three-mode and multi-set factor analysis using MRFA to estimate unique variances, and using CP, Tucker3 (for three-mode factor analysis), and Parafac2 (for muti-set factor analysis) to fit the covariance matrix of common part.

In Chapter 4, we did a simulation study that showed that three-mode factor analysis using CP performed very well in retrieving underlying factors when the data was randomly sampled from a normal distribution with a covariance matrix satisfying the CP covariance model. The solution of this model is unique up to permutation and scaling due to uniqueness properties of CP.

In Chapter 5, we extended the CP-based method in Chapter 4 to three-mode factor analysis using MRFA and Tucker3. Since solutions of this method are not unique due to non-uniqueness of Tucker3, a rotation method is needed. This extension was demonstrated by means of two applications to datasets in literature. The results of these two applications showed that one factor model is not always more appropriate than the other. In this chapter, we used the Joint Orhomax rotation in (Kiers, 1998a) to obtained simple structure of Tucker3 matrices.

In Chapter 6, we proposed and demonstrated an exploratory multi-set factor model with common covariance part of indirect Parafac2 form. The simulation study showed that our relatively simple procedure for the multi-set Parafac2 factor model performs very well in retrieving underlying factors when the data is randomly sampled with true covariance matrices satisfying the indirect Parafac2 model. The solution of this method is unique due to uniqueness properties of indirect Parafac2.

For our methods in Chapters 4, 5, 6, one can compute the percentage of

explained common variance which is usually not possible for other methods of three-mode or multi-set factor analysis. The reason for that is we use MRFA to estimate the unique variances. Moreover, the algorithms that we proposed are simple and easy to run. Our solutions are easy to interpret and our models are parsimonious.

In Chapter 7, we provided a classification of the three-mode and multi-set component and factor models featured in the thesis. Analogous to Tucker3-based three-mode factor model of Chapter 5, a Tucker3-based multi-set factor model exists that can be estimated analogous to the Parafac2 multi-set factor model of Chapter 6. This would be an interesting topic for future research.

Samenvatting (Summary in Dutch)

Dit proefschrift bespreekt exploratieve methoden voor componenten- en factoranalyse van meer-weg en meer-set data. In het bijzonder worden nieuwe methoden voor meer-weg en meer-set factoranalyse voorgesteld en gedemonstreerd. Nu volgt een samenvatting van de inhoud per hoofdstuk.

In hoofdstuk 2 bespreken we de achterliggende wiskundige concepten van meer-weg en meer-set componenten- en factormodellen. Deze zijn nodig om de volgende hoofdstukken volledig te kunnen begrijpen. Achtereenvolgens bespreken we twee-weg ontbindingen (e.g., de singuliere waarden ontbinding, principale componenten), drie-weg ontbindingen (e.g., Candecomp/Parafac (CP), Tucker3), meer-set componenten- en factormodellen (e.g., Parafac2, Simultaneous Component Analysis, meer-set Parafac2 common factor model). Verder bespreken we de uniciteitseigenschappen van zowel Candecomp/Parafac als Tucker3 en algoritmen om deze modellen te fitten.

In hoofdstuk 3 wordt een exploratieve analyse gedaan van vier-weg data afkomstig van een Belief in a Just World vragenlijst. De personen werd gevraagd hoe sterk zij geloven dat een aantal actoren (de natuur, god, menselijke instituties, andere mensen, jijzelf en het toeval) zorgt voor rechtvaardigheid in de wereld voor hemzelf of voor anderen. De vier 'wegen' in de dataset zijn: 345 personen, 7 items, 6 actoren en 2 perspectieven (voor jezelf en voor anderen). Eerst kijken we naar correlaties van items en actoren, naar t-toetsen om significante verschillen in scores op te sporen en naar principale componentenanalyse van twee-weg vormen van de dataset. Daarna fitten we Tucker3 en Tucker4 modellen. Vergeleken met de principale componentenanalyse, leveren de Tucker3 en Tucker4 modellen bijna hetzelfde resultaat. Ook hebben ze eenvoudig te interpreteren componenten. Echter, de Tucker3 en Tucker4 modellen bevatten veel

minder parameters dan de principale componentenanalyse. De Tucker4 oplossing bevat de minste parameters en wordt gefit op de complete vier-weg dataset. Aan het eind van het hoofdstuk bespreken we ook hoe de Tucker4 oplossing naar een eenvoudige vorm geroteerd kan worden.

In hoofdstukken 4, 5 en 6 introduceren en demonstreren we nieuwe modellen voor drie-weg en meer-set factoranalyse. We gebruiken Minimum Rank Factor Analysis (MRFA) om de unieke varianties te schatten en CP of Tucker3 (voor drie-weg factoranalyse) en Parafac2 (voor meer-set factoranalyse) voor het schatten van het gemeenschappelijke deel van de geobserveerde covarianties.

In hoofdstuk 4 doen we een simulatiestudie die laat zien dat drie-weg factoranalyse met MRFA en CP de onderliggende factoren goed schat wanneer de data normaal verdeeld is en de theoretische covariantiematrix aan het CP covariantiemodel voldoet. De oplossing van dit model is uniek tot op permutatie en schaling vanwege de uniciteitseigenschappen van CP.

In hoofdstuk 5 breiden we het CP covariantiemodel van hoofdstuk 4 uit naar drie-weg factoranalyse met MRFA en Tucker3. Omdat oplossingen van dit model niet uniek zijn (door de niet-uniciteit van Tucker3), is een rotatiemethode nodig. De uitbreiding naar Tucker3 wordt gedemonstreerd bij twee datasets uit de literatuur. De conclusie is dat het ene model niet altijd te prefereren is boven het andere model. Met andere woorden, beide modellen hebben bestaansrecht. In dit hoofdstuk gebruiken we de Joint Orthomax rotatiemethode van Kiers (1998a) voor de Tucker3 oplossingen.

In hoofdstuk 6 introduceren en demonstreren we een exploratief meer-set factormodel waarbij het gemeenschappelijke deel van de geobserveerde covarianties gemodelleerd wordt met behulp van het indirekte Parafac2 model. Een simulatiestudie laat zien dat onze relatief eenvoudige schattingsmethode de onderliggende factoren goed schat wanneer de data normaal verdeeld is en de the-

oretische covariantiematrices aan het Parafac2 covariantiemodel voldoen. De oplossing van dit factormodel is uniek vanwege de uniciteitseigenschappen van het indirekte Parafac2 model.

In onze factormodellen uit hoofdstukken 4, 5 en 6 kan een percentage verklaarde gemeenschappelijke variantie worden berekend, hetgeen voor andere modellen voor drie-weg of meer-set factoranalyse meestal niet mogelijk is. De reden hiervoor is dat we gebruik maken van MRFA voor het schatten van de unieke varianties. Verder zijn de algoritmen voor het schatten van onze factormodellen relatief eenvoudig en makkelijk te runnen. Onze oplossingen zijn makkelijk te interpreteren en onze modellen bevatten relatief weinig parameters.

In hoofdstuk 7 maken we een classificatie van de drie-weg en meer-set componenten- en factormodellen uit dit proefschrift. Analoog aan het Tucker3 covariantiemodel van hoofdstuk 5, bestaat er een Tucker3 meer-set covariantiemodel dat geschat kan worden analoog aan het Parafac2 meer-set covariantiemodel van hoofdstuk 6. Dit zou een interessant onderwerp kunnen zijn voor verder onderzoek.

Curriculum Vitae

Full name: LAM Thi Thanh Tam

Full name with accent in Vietnamese: Lâm Thị Thanh Tâm

LAM Thi Thanh Tam was born on 14 July 1980 in Ba To, Quang Ngai, Vietnam. She received her B.Sc. and M.Sc. degrees in Mathematics from Quy Nhon University, Vietnam, in 2002 and 2006, respectively. She has been a lecturer at the Department of Mathematics, Quy Nhon University, Vietnam since September 2002.

Since February 2011 she has been a Ph.D. student at Psychometrics and Statistics Group, Heymans Institute for Psychological Research, the Faculty of Behavioural and Social Sciences, the University of Groningen. Her Ph.D. project was funded by NWO (Dutch Scientific Research Organization) under the supervision of dr. drs. A.W. (Alwin) Stegeman.

Her Ph.D. research is in the field of Psychometrics and Statistics. Her main research interests are focused on multi-way and multi-set factor analysis.

Acknowledgements

The research presented in this thesis had been performed within Psychometrics and Statistics Group, Heymans Institute for Psychological Research, the Faculty of Behavioural and Social Sciences, the University of Groningen, the Netherlands in the past four years. This work has finished under the help, advice and contribution of my supervisor, my colleagues, and my friends. I would like to take an opportunity to thank them.

First and foremost, I am grateful and would like to thank Professor Rob Meijer and Associate Professor Alwin Stegeman for accepting me as their Ph.D. student at the Psychometrics and Statistics group and for supervising me for the past four years. As my promotor, Rob has given me tremendous freedom to work on my own, but he was always willing to give me advice and support whenever I needed. I truly feel so thankful him for everything he did for me. I would also like to express a deepest thanks to my supervisor, Alwin. He was everything I needed as a Ph.D. student, academically, and spiritually. I appreciate all his contribution of time, scientific ideas, many efforts, and kind heart to make my Ph.D. experience productive and stimulating. How much I would like to thank him for his thorough guidance, constructive discussion, positive remarks, and encouragement.

Next, I gratefully acknowledge the financial support of Dutch Scientific Research Organization (NWO) and the University of Groningen. Special thanks go to the secretaries of the Psychometrics and Statistics group for their kind support. Also, I am deeply grateful to the Department of Mathematics, Quy Nhon University, Vietnam for accepting my absence from work at there for a long period of time.

For this dissertation, I would also like to thank my reading committee mem-

bers for their time, interest, and constructive comments.

I would like to thank my colleagues with whom I shared so many things and from whom I received a lot of helps. My special thanks first go to my roommates Tanja Krone and Mariska Barendse for their helps, nice daily discussions, and for sharing the office. Many thanks also go to the colleagues and professors in the group of Psychometrics and Statistics. I would like to thank them for all interesting discussions and research presentations during our group seminars over the past years.

It is my pleasure to thank Katherine Stroebe of the Social Psychology Department, and my promotor Rob Meijer for providing us with their data. And I would also like to thank professor Henk A. L. Kiers for his helpful Matlab programs.

I would like to thank all my Vietnamese friends in Groningen who helped me, shared abroad living experience, and made my life in the Netherlands being less boring with unforgettable memories. Special thanks go to my dearest friends: Hao, Loan, Bay, Thang, Anh Nguyen, Tinh, Anh Pham, Hoa, Cuong-Huong for being by my side, encouraging, and supporting me.

Last but most of all, my greatest thanks must go to my dear family. I first would like to thank my parents, my brother and sister for their endless love and raising me. I would also like to thank my dear husband Quang Thuan and my dearest son Minh Khoi for all their love, always being my side, encouraging, support, and giving me power throughout these years.